amcs

KNOWLEDGE DISCOVERY IN DATA USING FORMAL CONCEPT ANALYSIS AND RANDOM PROJECTIONS

CHERUKURI ASWANI KUMAR

Networks and Information Security Division
School of Information Technology and Engineering, VIT University, Vellore, India
e-mail: `cherukuri@acm.org`

In this paper our objective is to propose a random projections based formal concept analysis for knowledge discovery in data. We demonstrate the implementation of the proposed method on two real world healthcare datasets. Formal Concept Analysis (FCA) is a mathematical framework that offers a conceptual knowledge representation through hierarchical conceptual structures called concept lattices. However, during the design of a concept lattice, complexity plays a major role.

Keywords: attribute implications, concept lattices, dimensionality reduction, formal concept analysis, knowledge discovery, random projections.

1. Introduction

Knowledge Discovery in Databases (KDD) is the non-trivial extraction of valid, implicit, potentially useful and understandable information from data. Recently, lattice theory under the framework of Formal Concept Analysis (FCA) has brought mathematical thinking for knowledge representation and discovery of knowledge (Venter *et al.*, 1997; Wille, 2001). FCA aims at discovering conceptual clusters (called formal concepts) in the data which are described in the form of formal context, discovering the data dependencies (called attributed implications), and visualizing them by a single conceptual structure called the concept lattice (Belohlavek and Vychodil, 2009). FCA organizes the information through concept lattices, which fundamentally comprises a partial order, modeling the subconcept–superconcept hierarchy (Jamil and Deogun, 2001). Unlike other partial order structures like trees, concept lattices allow multiple inheritances (Stumme, 2009). With their lattice structures and attribute implications, concept lattices support KDD. The crucial issue towards the discovery of knowledge using FCA is knowledge reduction while maintaining structure consistency (Aswani Kumar and Srinivas, 2010a; Belohlavek and Vychodil, 2010; Elloumi *et al.*, 2004, Wu *et al.*, 2009). Hence the goal is to minimize the input data before applying FCA. The literature has witnessed computationally expensive techniques like Singular Value De-

composition (SVD) (Aswani Kumar and Srinivas, 2010b; Pattison and Breiger, 2002; Snasel *et al.*, 2008).

Random Projections (RPs) have received significant attention from data mining and machine learning research communities for Dimensionality Reduction (DR) (Bingham and Mannila, 2001). RP devised under the conditions of orthonormality projects the given high dimensional data onto a lower dimensional subspace using a random matrix of unit length with normalized columns (Aswani Kumar, 2011). However, to the best of our knowledge, no analysis has been reported so far in the literature on RP based lattice reduction for a KDD task. The motivation behind the current paper is to make such an attempt by proposing an RP based lattice reduction method. The organization of the paper is as follows. Section 2 provides a brief background on FCA, RP and related work. Section 3 proposes an RP based FCA. Experimental results of the proposed method are reported in Section 4. Analysis and discussion of the results are presented in Section 5. Section 6 provides concluding remarks followed by acknowledgements and references.

2. Background

2.1. Formal concept analysis. Data analysis using FCA always starts with a formal context defined as a triple (**G, M,** *I*) where **G** is a set of formal objects, **M** is a set of formal attributes and *I* is the relation between **G** and **M**

(i.e., $I \subseteq \mathbf{G} \times \mathbf{M}$). A formal context is generally represented as a binary incidence table in which the crosses represent the binary relation between the object set and the attribute set. For the formal context, operators $\uparrow: 2^{\mathbf{G}} \rightarrow 2^{\mathbf{M}}$ and $\downarrow: 2^{\mathbf{M}} \rightarrow 2^{\mathbf{G}}$ are defined for every $\mathbf{A} \subseteq \mathbf{G}$ and $\mathbf{B} \subseteq \mathbf{M}$ by $A^{\uparrow} = \{m \in \mathbf{M}/\text{for each } g \in \mathbf{A} : \langle g, m \rangle \in I\}$, $B^{\downarrow} = \{g \in \mathbf{G}/\text{for each } m \in \mathbf{B} : \langle g, m \rangle \in I\}$. The operators $\uparrow$ and $\downarrow$ are known as concept forming operators. A formal concept of a formal context is defined as an ordered pair $(\mathbf{A}, \mathbf{B})$ with $\mathbf{A} \subseteq \mathbf{G}$ and $\mathbf{B} \subseteq \mathbf{M}$ such that $\mathbf{A}^{\uparrow} = \mathbf{B}$ and $\mathbf{B}^{\downarrow} = \mathbf{A}$. We call $\mathbf{A}$ and $\mathbf{B}$ respectively an extent and intent of the formal concept $(\mathbf{A}, \mathbf{B})$. A formal concept $(\mathbf{A}, \mathbf{B})$ of the context $(\mathbf{G}, \mathbf{M}, \ I)$ is defined to a be subconcept of the formal concept $(\mathbf{C}, \mathbf{D})$ of $(\mathbf{G}, \mathbf{M}, \ I)$ and $(\mathbf{C}, \mathbf{D})$ a superconcept of $(\mathbf{A}, \mathbf{B})$ if the extent $\mathbf{A}$ is contained in the extent $\mathbf{C}$ and, equivalently, if the intent $\mathbf{B}$ contains the intent $\mathbf{D}$. The set of all concepts of a context $(\mathbf{G}, \mathbf{M}, \ I)$ with the order relation $\leq$ is always a complete lattice called the concept lattice. Graphically represented concept lattices have proven to be useful in discovering and understanding conceptual relationships in given data (Wille, 2001).

The process of concept formation in FCA is considered a knowledge discovery task, and constructing the concept set constitutes the mining phase (Valtchev *et al.*, 2004; Poelmans *et al.*, 2010). There is a wealth of foundations available in the literature. For a detailed explanation of the basics of FCA and its mathematical properties, we refer the readers to a few authoritative references (Carpineto and Romano, 2004; Ganter and Wille, 1999; Stumme, 2009; Wille, 2008). Over the years, FCA has witnessed several applications in a wide variety of disciplines. Priss (2006) has given an exhaustive overview of FCA applications in information science. For requirements of real world scientific applications, different flavors of FCA have been proposed which generalize the classical concept lattice given by Ganter and Wille (1999). Currently, FCA has been extended to fuzzy concept lattices based on fuzzy theory, the rough formal concept based on rough set theory, the triadic concept, the monotone concept and the variable threshold concept lattice (Ghosh *et al.*, 2010).

FCA offers a compact mode of knowledge representation called attribute implications which are closely associated with functional dependencies in the database field and hence made their way into Association Rules Mining (ARM) in data mining (Aswani Kumar and Srinivas, 2010b). Approximative associations are implications with confidence less than 100% and exact associations are the valid implications with 100% confidence. The basis for the rules with 100% confidence is called the Duquenne–Guigues basis and the basis for the rules with confidence less than 100% is called the Luxenburger basis (Stumme, 2009). In contrast to other KDD techniques where queries or highly iterative approaches are followed, FCA allows graphical representation of the knowledge space that can be navigated by the user. The main inconvenience of FCA is the size of the formal context. The number H of nodes in a concept lattice has linear complexity with the number n of objects, $H = O(n \cdot 2^k)$, where k denotes the maximum number of attributes see the work of Aswani Kumar and Srinivas (2010a) and the references therein. The number of nodes and edges of a concept lattice can be reduced significantly by deriving equivalence relations between nodes of the lattice. Attribute implications are vulnerable to noise in the data. Scalability and computational tractability of FCA is another major concern in lattice based applications. Reducing the data to a lower dimension could help us in solving these problems.

Several interesting investigations are reported in the literature for reducing the concept lattice using different techniques including rough set theory by Liu *et al.* (2007), matrix rank reduction techniques like SVD and NMF by Snasel *et al.* (2008), Iceberg concept lattices by Stumme *et al.* (1998). Very recently, Aswani Kumar and Srinivas (2010b) have analyzed mining associations rules derived from SVD based reduced contexts. Though SVD provides the best approximation of the matrix with regard to the Frobenius norm, its computational complexity makes it impractical to apply over large size matrices (Aswani Kumar and Srinivas, 2006). For a dense matrix $\mathbf{X}$ of size $t \times d$, the complexity of computing SVD is $O(td^2)$, and for a sparse matrix with average c non-zero entries per data item the complexity is $O(tdc)$ (Aswani Kumar, 2009; Divya *et al.*, 2011). A very recent interesting investigation by Aswani Kumar and Srinivas (2010a) shows that FKM clustering can successfully be applied for concept lattice reduction.

2.2. Random projections. In RP, DR can be achieved by projecting the data matrix through the origin onto a lower dimensional subspace formed by a set of random vectors. Thus, given a data matrix $\mathbf{X}$ of size $t \times d$, where d is the original dimensionality of data and t is the total number of points, the dimensionality of $\mathbf{X}$ can be reduced by projecting it onto a k dimensional subspace $(k < d)$ as

$$\mathbf{D} = \mathbf{X}_{t \times d} \mathbf{R}_{d \times k}, \tag{1}$$

where $\mathbb{R}_{d \times k}$ is a random projection matrix (Bingham and Mannila, 2001). The RP idea stems from the Johnson–Lindenstrauss lemma, which states that, if points in a vector space are randomly projected onto a lower dimensional subspace, $\mathbb{R}^d \rightarrow \mathbb{R}^k$, then the Euclidean distance among the set of n points is preserved for any $0 < \epsilon < 1$.

Theorem 1. (Johnson–Lindenstrauss lemma) *For any* $0 < \epsilon < 1$ *and any integer* n, *let* k *be a positive integer such that*

$$k \geq 4\big[(\epsilon^2/2) - (\epsilon^3/3)\big]^{-1} \ln n. \tag{2}$$

Then for any set **W** of n points in $\mathbb{R}^d$ there is a map f, $f : \mathbb{R}^d \to \mathbb{R}^k$, such that, for all $u, v \in W$,

$$(1 - \epsilon) \|u - v\|^2 \le \|f(u) - f(v)\|^2$$
$$\le (1 + \epsilon) \|u - v\|^2. \qquad (3)$$

Further this map f can be found in randomized polynomial time (Achilioptas, 2003). The choice of random matrix **R** is a crucial design decision and different approaches are suggested for the generation of random matrices. The elements of **R** are often Gaussian distributed. Unlike other matrix decomposition methods like SVD, RP does not require computationally expensive calculations like eigenvalue decompositions. Several applications of RP such as information retrieval, handwritten text recognition, image compression, face recognition, etc. are reported in the literature (Aswani Kumar, 2011; Varmuza *et al.*, 2010).

3. Proposed method

In this section we propose to apply an RP based FCA method for KDD. Then we apply FCA on the reduced context for a knowledge discovery task. The following is the proposed method:

1. From the given formal context, construct the binary object-attribute incidence matrix of size $t \times d$ with a dimensionality d. Normalize the columns of the data matrix to unit length.

2. Choose a lower dimension value $k \, (k \le d)$. Create a random matrix **R** of size $d \times p$.

3. Apply an RP over the original matrix to obtain the reduced dimensionality matrix **D** as given in Eqn. (1).

4. Compute the reduced dimensionality matrix $\mathbf{D}_k$ of the original data matrix **X** as $\mathbf{D}_k = \mathbf{DZ}^*$, where $\mathbf{Z}^*$ is given as $\mathbf{Z}^* = (\mathbf{D}^T \mathbf{D}^{-1} \mathbf{D}^T \mathbf{X}$.

5. The reduced matrix $\mathbf{D}_k$ does not contain binary values. Hence choose a threshold value $t, 0 \le t \le 1$, for discriminating between 1s and 0s. Then construct the reduced rank binary matrix.

6. Construct the reduced formal context from the reduced binary incidence matrix. Apply FCA to the reduced context.

Thus with the proposed method, the formal context reduces to a lower dimension which can lead to minimizing the input data to handle the scalability and computational tractability of FCA. The entire operation of the RP on a data matrix of size $t \times d$ to a reduced dimensionality k is $O(tdk)$ (cf. Achilioptas, 2003; Aswani Kumar, 2010).

4. Experimental results

In this section we report experimental results of the proposed method on two healthcare datasets. These datasets are obtained from a healthcare informatics project of the Medical Research Council of South Africa (Horner, 2007). The diseases studied in the project are tuberculosis (TB), chronic bronchitis (CB) and hypertension (HP). However, in our analysis we focus on TB and HP datasets due to the fact that the CB dataset does not contain test data. Experiments on TB and HP datasets are conducted in three phases. In the first phase, we applied FCA and RP based FCA over the training dataset. In the second phase, we tested the quality of the derived knowledge, described in the form of rules, with domain experts' knowledge and a test dataset. In the third stage, we compared RP-FCA with SVD-FCA. Performance is measured by identifying the number of times that rules from FCA, RP-FCA and SVD-FCA have the same conclusion as the treating doctor.

4.1. Experiments on tuberculosis data. The tuberculosis dataset contains details of 21 patients for 12 symptoms related to TB. Table 1 lists various TB symptoms. Table 2 lists domain experts' knowledge, in the form of rules, in determining TB using the symptoms listed in Table 1. Table 3 shows the formal context. The last column of the matrix indicates the presence or absence of TB. Figure 1 shows the concept lattice obtained by applying FCA over the TB context given in Table 3. Each node of the lattice structure represents a concept. The concept lattice shown in Fig. 1 is of height 11 with 101 concepts and 253 edges. Along with the concept lattice, FCA also produced 33 implications in the DG basis. However, implications which describe TB are of interest in this study. Table 4 lists all such implications.

Table 1. TB symptoms.

No.	Symptom	Abbreviation
1	Persistent cough	PC
2	Sputum production	SP
3	Sputum produced is muco-purulent	MC
4	Bloody sputum	BS
5	Clear sputum	CS
6	Weight loss	WL
7	Extreme night sweats	NS
8	No appetite	NA
9	Chest pain	CP
10	Shortness of breath	SB
11	Tuberculosis contact	TC
12	Tiredness	TN

If the antecedent of an implication which has the target attribute in its consequent is a subset of the antecedent of an expert rule, then we can consider the expert rule to be subsumed by the implication. It is clear from Table 4

that the antecedents of implications 4–9 are subsets of the antecedents of expert rules. Implications 1–3 are not part of expert rules. None of the implications in the DG basis is overlapped exactly with any of the expert rules. From Table 4 we can also observe that implications 4–9 make the diagnosis of the disease with fewer symptoms. Table 5 shows how expert rules are subsumed using implications from FCA and RP-FCA with different values of k and t, which we will explain later.

Table 2. Expert rules for TB.

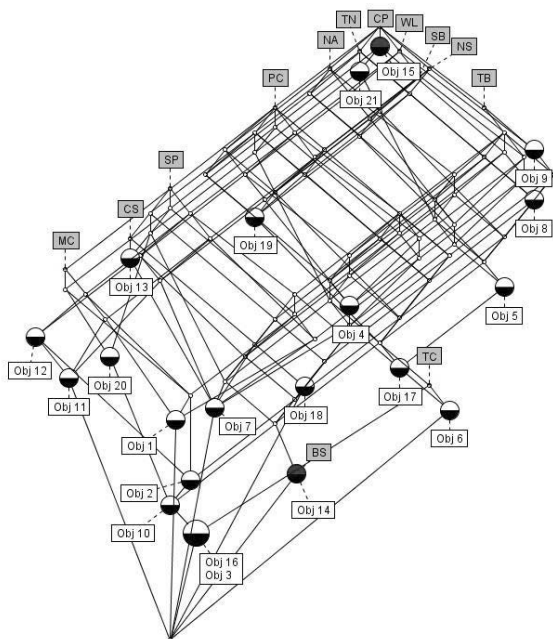| Sl. no	Expert rules for tuberculosis
1	PC SP BS WL $\longrightarrow$ TB
2	PC SP BS NS $\longrightarrow$ TB
3	PC WL NS $\longrightarrow$ TB
4	PC SP BS TC $\longrightarrow$ TB
5	PC SP BS CP NA $\longrightarrow$ TB
6	PC SP BS SB $\longrightarrow$ TB
7	PC WL CP SB $\longrightarrow$ TB
8	PC SP BS CP TN $\longrightarrow$ TB

Fig. 1. Concept lattice of the TB training context.

Next we applied the proposed RP based FCA on the TB training incidence matrix, which is of size 21×13 with rank 12. RP with $k = 6$ and a binary threshold $t = 0.5$ is applied over the incidence matrix. An entry in the reduced matrix is considered to be 1 if it is higher than the chosen t value, and 0 otherwise. We applied FCA over the reduced context. Figure 2 shows the concept lattice containing a concept count of 67 with 152 edges and a height of 9. RP-FCA with $k = 6$ and $t = 0.5$ produced 28 implications in the DG basis. Implications concluding TB are shown

Table 4. Implications obtained from TB training data using FCA.

No.	Implication	No. of objects implication holds
	Implications NOT part of expert rules	
1	NS CP$\rightarrow$ TB	8
2	WL TN $\rightarrow$ TB	11
3	PC SP MC CS CP TN $\rightarrow$ TB	0
	Implications part of expert rules	
4	NA CP $\rightarrow$ TB	9
5	BS $\rightarrow$ TB	1
6	PC SP NS $\rightarrow$TB	6
7	WL NS $\rightarrow$ TB	10
8	WL CP $\rightarrow$ TB	10
9	TC $\rightarrow$ TB	3

in Table 6. New implications 4–8 are part of expert rules. Hence they are used in subsuming all the expert rules as shown in Table 5. We also studied the influence of discretization parameter t by choosing $t = 0.8$. FCA on the reduced context with $k = 6$ and $t = 0.8$ produced a concept lattice of height 8 with 51 concepts and 115 edges as shown in Fig. 3. New implications obtained from this analysis are shown in Table 7. It is clear that implications 4–12 are part of expert rules and hence they are used in subsuming expert rules as shown in Table 5.

For further analysis we conducted experiments using different values of k and t. FCA on the reduced context with $k = 3$ and $t = 0.5$ produced 18 implications in DG basis. Table 8 lists 8 implications inferring TB. Implications 6–8 are part of expert rules and hence they are used in subsuming expert rules as shown in Table 5. It is interesting to note that none of the newly produced implications subsumed the expert rule PC WL CP SB $\rightarrow$ TB. The lattice structure obtained from this analysis has 42 concepts with 80 edges and a height of 10.

Table 6. Implications obtained from TB training data with RP $k = 6$ and $t = 0.5$.

Sl. no.	Implication	No. of objects for which the implication holds
	Implications NOT part of expert rules	
1	NS CP$\rightarrow$ TB	8
2	WL TN $\rightarrow$ TB	13
3	CS $\rightarrow$ TB	2
	Implications part of expert rules	
4	NA CP $\rightarrow$ TB	9
5	BS $\rightarrow$ TB	0
6	WL NS $\rightarrow$ TB	11
7	WL CP $\rightarrow$ TB	10
8	TC $\rightarrow$ TB	1

FCA on the RP based reduced context with $k = 3$ and $t = 0.8$ produced a lattice structure of height 8 with 42 concepts and 88 edges. FCA also produced 19 implications in the DG basis, among which 10 implications

Table 3. Incidence matrix of the TB training dataset.

	PC	SP	MC	BS	CS	WL	NS	NA	CP	SB	TC	TN	TB
Obj 1	×	×	×			×			×	×			×
Obj 2	×	×	×			×	×	×	×				×
Obj 3	×	×	×			×	×	×	×		×	×	×
Obj 4	×					×		×		×		×	×
Obj 5						×	×	×	×	×		×	×
Obj 6	×					×	×	×	×	×	×	×	×
Obj 7	×	×			×	×		×	×	×		×	×
Obj 8						×	×					×	×
Obj 9							×						×
Obj 10	×	×	×			×	×	×	×			×	×
Obj 11	×	×	×			×		×		×			
Obj 12	×	×	×			×		×					
Obj 13	×	×			×				×			×	
Obj 14	×	×		×		×	×	×	×			×	×
Obj 15									×				
Obj 16	×	×	×			×	×	×	×		×	×	×
Obj 17	×					×	×	×	×	×		×	×
Obj 18	×	×			×	×	×					×	×
Obj 19	×						×	×				×	
Obj 20	×	×	×						×			×	
Obj 21									×			×	

Table 5. Subsuming expert rules.

Expert rule number	Implication number in FCA	Implication number in RP-FCA $k = 6, t = 0.5$	Implication number in RP-FCA $k = 6, t = 0.8$	Implication number in RP-FCA $k = 3, t = 0.5$	Implication number in RP-FCA $k = 3, t = 0.8$
1	5	5	4	6	5
2	6	5	4	6	5, 7
3	7	6	5,7	7	7
4	5, 9	5, 8	4,11	6,8	5, 6
5	4, 5, 9	5	4	6	5, 8
6	5, 9	5	4,10	6	5, 9
7	8	7	9, 10	–	8, 9
8	5	5	4, 12	6	5, 8

infer TB as shown in Table 9. Implications 5–10 are used in subsuming expert rules as shown in Table 5. We compared the performance of RP based FCA with SVD based FCA (Aswani Kumar and Srinivas, 2010b). Table 10 summarizes the number of concepts, edges, implications in the DG basis, new rules, the height of the concept lattice and the number of expert rules subsumed using FCA, RP-FCA and SVD-FCA. From this summary we can understand that, even with fewer concepts, RP-FCA, except ($k = 3$, $t = 0.8$), is able to subsume all the expert rules and produce new implications similarly to SVD-FCA.

The next step of our analysis is to verify the quality of the new implications obtained using RP-FCA and compare them with that of SVD-FCA. Table 11 shows the test dataset which contains symptoms of 10 patients and treating doctors' conclusion on the presence of TB for each of

the patient. Expert rules and implications produced from FCA, RP-FCA and SVD-FCA with different values of k and t are compared on the test dataset, and their performance results are summarized in Table 12. From Table 11 we can identify that, among 10 patients, treating doctors confirmed TB for 9 patients (23 to 31). However, expert rules confirmed the disease for 7 patients and did not confirm the disease for 1 patient (Patient 22). Expert rules failed to confirm TB for Patients 30 and 31. Hence the diagnosis accuracy of expert rules is 80%.

Next we analyzed the implications obtained by applying FCA, RP-FCA and SVD-FCA. From Table 12 we can understand that FCA, RP-FCA (excluding $k = 3$, $t = 0.8$) and SVD-FCA correctly identified the presence or absence of TB for 9 patients. Patient 30 has only one symptom—weight loss. However, for this patient, doctors' diagnosis revealed TB. Rules produced from FCA,

Table 10. Summary of FCA results on TB data.

	Concepts	Edges	No. of implications in the DG basis	Height of the lattice	No. of new implications	No. of expert rules subsumed
FCA	101	253	33	11	9	8
RP-FCA $k = 6$, $t = 0.5$	67	152	28	9	8	8
RP-FCA $k = 6$, $t = 0.8$	51	115	21	8	12	8
RP-FCA $k = 3$, $t = 0.5$	42	80	18	10	8	7
RP-FCA $k = 3$, $t = 0.8$	42	88	19	8	10	8
SVD-FCA $k = 6$, $t = 0.5$	98	247	31	11	9	8
SVD-FCA $k = 6$, $t = 0.3$	74	177	28	8	8	8
SVD-FCA $k = 3$, $t = 0.5$	44	89	18	9	6	8
SVD-FCA $k = 3$, $t = 0.8$	32	62	19	7	9	7

Table 11. Incidence matrix of TB test data.

	PC	SP	MC	BS	CS	WL	NS	NA	CP	SB	TC	TN	TB
Obj 22	×								×			×	
Obj 23	×	×	×			×	×		×	×		×	×
Obj 24	×	×	×			×	×	×	×		×	×	×
Obj 25	×	×	×			×		×	×	×	×	×	×
Obj 26	×	×	×			×	×	×			×	×	×
Obj 27	×	×	×			×	×		×	×	×	×	×
Obj 28	×	×			×	×	×	×				×	×
Obj 29	×					×	×		×			×	×
Obj 30						×							×
Obj 31						×	×	×					×

Table 12. Performance comparisons of FCA, RP-FCA and SVD-FCA on TB test data.

Patient	Doctors' assess-ment	Expert rules	FCA	RP-FCA $k = 6$, $t = 0.5$	RP-FCA $k = 6$, $t = 0.8$	RP-FCA $k = 3$, $t = 0.5$	RP-FCA $k = 3$, $t = 0.8$	SVD-FCA $k = 6$, $t = 0.5$	SVD-FCA $k = 6$, $t = 0.8$
Obj 22	–	–	–	–	–	–	TB	–	–
Obj 23	TB	TB	TB	TB	TB	TB	TB	TB	TB
Obj 24	TB	TB	TB	TB	TB	TB	TB	TB	TB
Obj 25	TB	TB	TB	TB	TB	TB	TB	TB	TB
Obj 26	TB	TB	TB	TB	TB	TB	TB	TB	TB
Obj 27	TB	TB	TB	TB	TB	TB	TB	TB	TB
Obj 28	TB	TB	TB	TB	TB	TB	TB	TB	TB
Obj 29	TB	TB	TB	TB	TB	TB	TB	TB	TB
Obj 30	TB	–	–	–	–	–	–	–	–
Obj 31	TB	–	TB	TB	TB	TB	TB	TB	TB
Performance		80%	90%	90%	90%	90%	80%	90%	90%

Table 14. Incidence matrix of HP training data.

	OH	BV	DZ	NU	NB	OW	FH	LR	HP
Obj 1	×	×	×	×					×
Obj 2	×	×	×		×			×	×
Obj 3	×		×			×	×	×	×
Obj 4	×		×				×	×	×
Obj 5	×		×			×	×	×	×
Obj 6	×	×	×	×	×			×	×
Obj 7	×		X				×	×	×
Obj 8	×	×	×	×		×		×	×
Obj 9			×				×	×	×
Obj 10			×		×		×	×	×
Obj 11	×	×					×	×	

Table 7. Implications obtained from TB training data with RP $k = 6$ and $t = 0.8$.

Sl. no.	Implication	No. of objects for which the implication holds
	Implications NOT part of expert rules	
1	CS → TB	0
2	NS CP → TB	6
3	WL TN → TB	4
	Implications part of expert rules	
4	BS → TB	0
5	PC NS → TB	4
6	SP NS → TB	4
7	WL NS → TB	7
8	NA CP → TB	6
9	WL CP → TB	7
10	SB → TB	1
11	TC → TB	0
12	SP TN → TB	2

Table 8. Implications obtained from TB training data with RP $k = 3$ and $t = 0.5$.

Sl. no.	Implication	No. of objects for which the implication holds
	Implications NOT part of expert rules	
1	CS → TB	0
2	NS CP → TB	10
3	WL NA TN → TB	11
4	NS TN → TB	12
5	PC WL CP SB TN → TB	4
	Implications part of expert rules	
6	BS → TB	0
7	WL NS → TB	11
8	TC → TB	2

Table 9. Implications obtained from TB training data with RP $k = 3$ and $t = 0.8$.

Sl. no.	Implication	No. of objects implication holds
	Implications NOT part of expert rules	
1	CS → TB	0
2	WL TN → TB	6
3	NA TN → TB	6
4	PC SP MC WL NA → TB	2
	Implications part of expert rules	
5	BS → TB	0
6	TC → TB	0
7	NS → TB	6
8	CP → TB	6
9	SB → TB	0
10	SP TN → TB	2

RP-FCA and SVD-FCA failed to confirm the disease for this patient. Implication from RP-FCA ($k = 3$, $t = 0.8$) confirmed TB for Patient 22 wrongly. From this analysis we can infer that, with the reduced context, implications from RP-FCA are able to subsume all the expert rules and
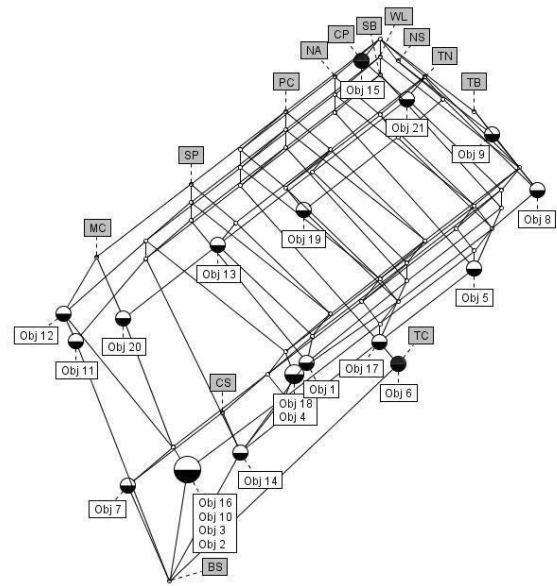

Fig. 2. Concept lattice of the TB training context with RP $k = 6$, $t = 0.5$.
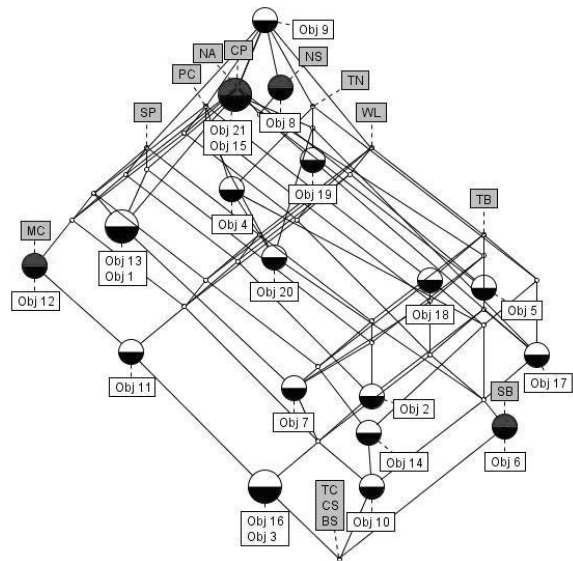

Fig. 3. Concept lattice of the TB training context with RP $k = 6$, $t = 0.8$.

diagnose TB better than expert rules and similarly to FCA and SVD-FCA.

4.2. Experiments on hypertension data. The hypertension dataset includes training data of 11 patients and testing data of 10 patients for various symptoms of HP listed in Table 13. However, expert rules are not available for this disease. Table 14 shows the formal context of HP training data. Figure 4 shows the concept lattice structure obtained by applying FCA on the training context.

The lattice structure is of height 7 and contains 24 concepts with 42 edges. FCA also produced 11 implications in the DG basis, out of which we considered 4 impli-

Table 13. HP symptoms.

No.	Symptom	Abbreviation
1	Persistent occipital headache	OH
2	Blurred vision	BV
3	Dizziness	DZ
4	Nausea	NU
5	Nose bleed	NB
6	Overweight	OW
7	Family hypertension	FH
8	Lifestyle risk	LR

Table 15. Implications from HP training data.

Sl. no.	Implication	No. of objects implication holds
implications from FCA on HP training data		
1	DZ → HP	10
2	NU → HP	3
3	NB → HP	3
4	OW → HP	3
Implications from RP-FCA $k = 3$, $t = 0.5$		
1	BV → HP	4
2	DZ → HP	9
3	NB → HP	3
4	NU → HP	2
Implications from RP-FCA $k = 3$, $t = 0.8$		
1	OH → HP	5
2	BV → HP	1
3	DZ → HP	6
4	NU → HP	2
5	NB → HP	0
6	OW → HP	1

cations, listed in Table 15, that infer HP. The HP training binary incidence matrix shown in Table 14 is of size 11×9 with rank 9. We applied RP on this matrix choosing the number of projections k to be 3 and the binary threshold limit t to be 0.5 and obtained a reduced context. FCA on this reduced context produced a lattice structure, shown in Fig. 5, having a height of 7 and containing 17 concepts with 26 edges.
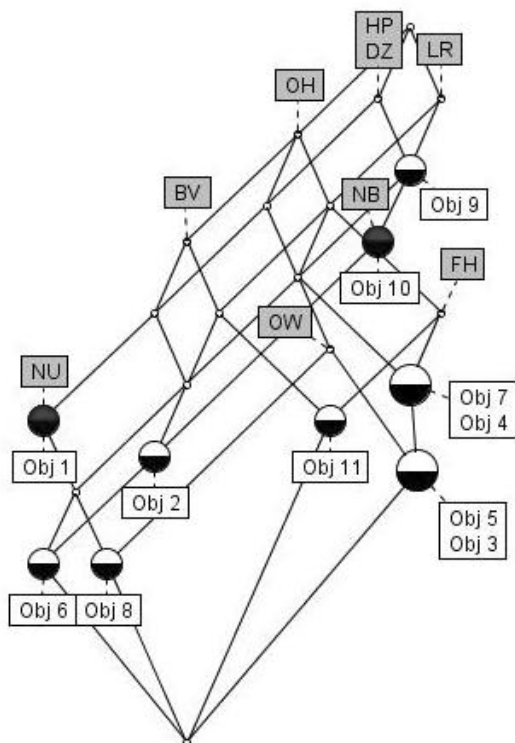

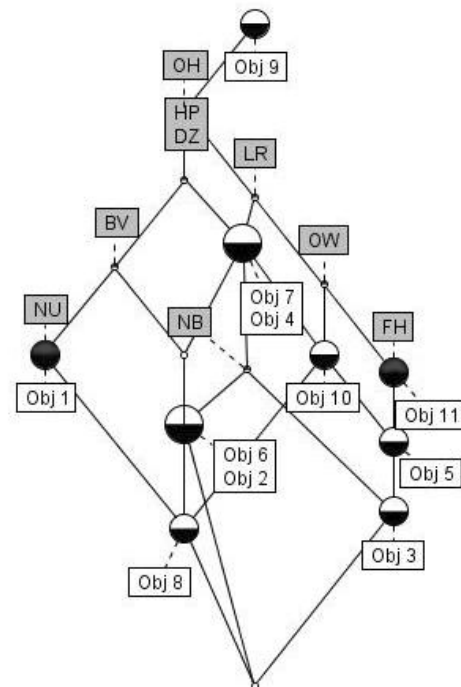Fig. 4. Concept lattice of the HP training context.

Fig. 5. Concept lattice of the HP training context with RP $k = 3$ $t = 0.5$.

FCA also produced 12 implications in the DG basis out of which 4 implications listed in Table 15 infer HP. For further analysis we applied RP with $k = 3$ and $t = 0.8$ on the HP training binary incidence matrix. Six implications that infer HP are listed in Table 15. For comparative study we applied SVD on the HP training matrix with different values the number of projections (k) and binary threshold limits (t). Table 16 summarizes the number of concepts,

edges, the height of lattice structures, the number of implications in the DG basis and the number of new implications that make positive conclusions about HP, obtained by applying FCA on the original and reduced contexts.

The next step of our analysis is to verify the quality of the new knowledge in the form of rules produced from FCA, RP-FCA and SVD-FCA based reduced contexts. Table 17 lists the binary formal context of HP test

Table 16. Summary of the implications from HP data.

	Concepts	Edges	Implications in the DG basis	Height of the lattice	No. of new rules
FCA	24	42	11	7	4
RP-FCA $k = 3$, $t = 0.5$	17	26	12	7	4
RP-FCA $k = 3$, $t = 0.8$	8	10	11	4	6
SVD-FCA $k = 3$, $t = 0.5$	13	19	11	5	5
SVD-FCA $k = 3$, $t = 0.8$	12	16	9	6	5

Table 17. Incidence matrix of HP test data.

	OH	BV	DZ	NU	NB	OW	FH	LR	HP
Obj 12	×				×	×	×	×	×
Obj 13	×	×	×					×	×
Obj 14						×	×		×
Obj 15	×	×	×				×	×	×
Obj 16	×		×	×					×
Obj 17	×	×	×	×		×	×	×	×
Obj 18			×			×	×		×
Obj 19									×
Obj 20	×		×						
Obj 21	×	×					×	×	×

Table 18. Performance comparisons of FCA and SVD-FCA on HP test data.

Patient	Doctors' assessment	FCA	RP-FCA $k = 3$, $t = 0.5$	RP-FCA $k = 3$, $t = 0.8$	SVD-FCA $k = 3$, $t = 0.5$	SVD-FCA $k = 3$, $t = 0.8$
Obj 12	HP	HP	HP	HP	HP	HP
Obj 13	HP	HP	HP	HP	HP	HP
Obj 14	HP	HP	–	HP	HP	HP
Obj 15	HP	HP	HP	HP	HP	HP
Obj 16	HP	HP	HP	HP	HP	HP
Obj 17	HP	HP	HP	HP	HP	HP
Obj 18	HP	HP	HP	HP	HP	HP
Obj 19	HP	–	–	–	–	–
Obj 20	–	HP	HP	HP	HP	HP
Obj 21	HP	–	HP	HP	HP	HP
Performance		70%	70%	80%	80%	80%

Table 19. TB and HP datasets results.

	True Positive (TP)	True Negative (TN)	False Positive (FP)	False Negative (FN)
Outcomes from the TB dataset				
Expert rules	7	1	0	2
FCA	8	1	0	1
RP-FCA $k = 6$, $t = 0.5$	8	1	0	1
RP-FCA $k = 6$, $t = 0.8$	8	1	0	1
RP-FCA $k = 3$, $t = 0.5$	8	1	0	1
RP-FCA $k = 3$, $t = 0.8$	8	0	1	1
SVD-FCA $k = 6$, $t = 0.5$	8	1	0	1
SVD-FCA $k = 6$, $t = 0.8$	8	1	0	1
Outcomes from the HP dataset				
FCA	7	0	1	2
RP-FCA $k = 3$, $t = 0.5$	7	0	1	2
RP-FCA $k = 3$, $t = 0.8$	8	0	1	1
SVD-FCA $k = 3$, $t = 0.5$	8	0	1	1
SVD-FCA $k = 3$, $t = 0.8$	8	0	1	1

Table 20. Specificity and sensitivity analysis of results on TB data.

	Expert rules	FCA	RP-FCA $k = 6$, $t = 0.5$	RP-FCA $k = 6$, $t = 0.8$	RP-FCA $k = 3$, $t = 0.5$	RP-FCA $k = 3$, $t = 0.8$	SVD-FCA $k = 6$, $t = 0.5$	SVD-FCA $k = 6$, $t = 0.8$
Specificity	100%	100%	100%	100%	100%	0%	100%	100%
Sensitivity	78%	89%	89%	89%	89%	89%	89%	89%
No. of new implications	–	3	3	3	5	4	4	4

Table 21. Specificity and sensitivity analysis of results on HP data.

	FCA	RP-FCA $k = 3$, $t = 0.5$	RP-FCA $k = 3$, $t = 0.8$	SVD-FCA $k = 3$, $t = 0.5$	SVD-FCA $k = 3$, $t = 0.8$
Specificity	0%	0%	0%	0%	0%
Sensitivity	78%	78%	89%	89%	89%

data containing different symptoms and treating doctors' conclusion about the presence or absence of HP for a set of 10 patients. It is clear from Table 17 that, out of 10 patients, 9 have the disease, HP. Table 18 summarizes the performance of new implications derived from FCA, RP-FCA and SVD-FCA. It is clear from Table 18 that implications produced from FCA correctly diagnosed the presence or absence of the disease for 7 patients. Rules discovered from RP-FCA ($k = 3$, $t = 0.8$) produced 80% accuracy, similarly to SVD-FCA. We can observe from Table 18 that RP-FCA ($k = 3$, $t = 0.8$) failed to give correct diagnosis for Patients 19 and 20. Though treating doctors concluded the presence of HP for Patient 19, implications from FCA, RP-FCA and SVD-FCA failed to confirm the presence of HP due to the fact that the patient does not have any symptoms listed in Table 13.

An interesting result can be observed for Patient 20 who has persistent occipital headache and dizziness symptoms, and doctors' diagnosis did not conclude the disease. However, implications from FCA on the original context confirmed the disease for Patient 20. Though this False Positive (FP) result remains with RP-FCA ($k = 3$, $t = 0.8$), the number of False Negative (FN) outcomes decreased, similarly as in SVD-FCA. From Table 18 we can also observe that implications from FCA on the original context fail to detect HP for Patient 21. Implications derived from RP-FCA and SVD-FCA confirmed the disease. It is interesting to observe that RP-FCA ($k = 3$, $t = 0.5$) correctly identified the presence or absence of the disease for 7 patients. Similarly to FCA, RP-FCA ($k = 3$, $t = 0.5$) produced one false positive and two false negative conclusions. To conclude, FCA on the RP based reduced context with $k = 3$, $t = 0.8$ performed better than FCA and similarly to FCA on the SVD based HP reduced context.

5. Analysis

This paper has proposed an RP based FCA for a KDD task. Experiments are conducted on healthcare data using Matlab 6.5 and ConExp (http://conexp.sourceforge.net/). During the entire analysis, our focus was on the implications with 100% confidence and support of zero due to the fact that the data are related to healthcare. A low support implication in DG basis can still be valid if it does not contradict any example of the context. Though the entire analysis of FCA, RP-FCA on TB data produced the implication BS → TB with low support, it was observed by treating doctors that bloody sputum is a very serious symptom of TB. The analysis may result in false positive outcomes when the implications with low support are used. However in the case of healthcare, the ethical challenge is that false negative outcomes are of greater risk than FP outcomes. Table 19 summarizes the number of True Positive (TP), True Negative (TN), false positive and false negative outcomes obtained on TB and HP datasets using different methods. It is clear from Table 19 that, on the HP data, the number of FN outcomes is decreased and that of TP outcomes is increased by using implications from RP-FCA ($k = 3$, $t = 0.8$). On the TB dataset it is clear that RP-FCA performed similarly to FCA and SVD-FCA.

For further investigations we tested the specificity and sensitivity of the results obtained in Table 19. Specificity measures the proportion of the TNs which are correctly classified. Sensitivity indicates the proportion of the TPs which are correctly classified. Tables 20 and 21 present the specificity and sensitivity of the results shown in Table 19. Specificity analysis on the TB test data results indicate that expert rules and implications from FCA, RP-FCA (except $k = 3$, $t = 0.8$) and SVD-FCA are suc-

cessful in diagnosing all healthy people as healthy. Since implications from RP-FCA ($k = 3$, $t = 0.8$) did not correctly diagnose the only available healthy patient (Patient 22), its specificity is zero. However, on HP test data, the specificity of all methods is zero since Patient 20 is the only healthy person and implications from FCA, RP-FCA and SVD-FCA wrongly diagnosed him as an HP patient. Sensitivity analysis on TB test data shows that RP-FCA identified more TPs than expert rules, similarly to FCA and SVD-FCA. Sensitivity analysis on HP test data indicates that RP-FCA ($k = 3$, $t = 0.8$) identified more TPs than FCA, similarly to SVD-FCA.

From Table 10 it is clear that, on the TB dataset, the numbers of implications generated by FCA, RP-FCA ($k = 6$, $t = 0.8$; $k = 3$, $t = 0.8$), SVD-FCA ($k = 6$, $t = 0.5$; $k = 3$, $t = 0.8$) are greater than the number of rules given by experts. From Table 16 we can observe that, on HP data, RP-FCA ($k = 3$, $t = 0.8$) and SVD-FCA produced more implications than FCA. From the implication tables we can understand that all the implications produced from RP-FCA used a smaller number of symptoms for diagnosis while expert rules about TB used more symptoms. On the TB dataset, implications produced by RP-FCA (except $k = 3$, $t = 0.5$) subsumed all the expert rules. This can be considered another significant result as the rules suggested by experts are part of accepted medical knowledge. Table 10 lists also the number of new implications produced by FCA, RP-FCA and SVD-FCA on the TB dataset. None of these implications match expert rules, hence they are considered new knowledge about the disease. However, on the HP dataset, since no expert rules are available, all the implications produced by FCA, RP-FCA and SVD-FCA are considered new knowledge. Future work may concentrate on mining indirect associations using FCA (Kazienko, 2009) and mining associations with background knowledge and priorities of the attributes (Belohlavek and Vychodil, 2009) with other reduction methods (Aswani Kumar, 2009; Aswani Kumar and Srinivas, 2010c).

6. Conclusions

A central issue in FCA based KDD is the size of the formal context. To address this issue, we proposed an RP based FCA. Experiments were conducted on two real world healthcare datasets. Knowledge derived in the form of attribute implications from the RP based FCA reduced context was analyzed and compared with that of SVD based FCA. A summary of the findings is as follows:

- RP based FCA is computationally less costly than SVD based FCA.

- Knowledge derived from the RP-FCA based reduced TB context is able to subsume all the expert rules

and diagnose TB similarly to FCA and SVD-FCA but better than expert rules.

- Knowledge derived from the RP-FCA based reduced HP context ($k = 3$, $t = 0.8$) diagnosed the disease similarly to SVD-FCA but better than FCA.

- Specificity and sensitivity analysis on the results obtained confirmed the above observations.

References

Achilioptas, D. (2003). Database friendly random projections: Johnson–Lindenstrauss with binary coins, *Journal of Computer and System Sciences* **66**(4): 671–687.

Aswani Kumar, Ch. (2009). Analysis of unsupervised dimensionality reductions, *Computer Science and Information Systems* **6**(2): 217–227.

Aswani Kumar, Ch. (2010). Random projections for concept lattice reduction, *Proceedings of the 4th International Conference on Information Processing, Bengaluru, India,* pp. 1-11.

Aswani Kumar, Ch. (2011). Reducing data dimensionality using random projections and fuzzy k-means clustering, *International Journal of Intelligent Computing and Cybernetics* **4**(3): 353–365.

Aswani Kumar, Ch. and Srinivas, S. (2006). Latent semantic indexing using eigenvalue analysis for efficient information retrieval, *International Journal of Applied Mathematics and Computer Science* **16**(4): 551–558.

Aswani Kumar, Ch. and Srinivas, S. (2010a). Concept lattice reduction using fuzzy k-means clustering, *Expert Systems with Applications* **9**(1): 2696-2704.

Aswani Kumar, Ch. and Srinivas, S. (2010b). Mining associations in health care data using formal concept analysis and singular value decomposition, *Biological Systems* **18**(4): 787–807.

Aswani Kumar, Ch. and Srinivas, S. (2010c). A note on weighted fuzzy k-means clustering for concept decomposition, *Cybernetics and Systems* **41**(6): 455–467.

Belohlavek, R. and Vychodil, V. (2009). Formal concept analysis with background knowledge: Attribute priorities, *IEEE Transactions on Systems, Man and Cybernetics* **39**(4): 399–409.

Belohlavek, R. and Vychodil, V. (2010). Discovery of optimal factors in binary data via a novel method of matrix decomposition, *Journal of Computer and System Sciences* **76**(1): 3–20.

Bingham, E. and Mannila, H. (2001). Random projections in dimensionality reduction, *Proceedings of the 7th International Conference on Knowledge Discovery and Data Mining, San Franscisco, CA, USA*, pp. 245–250.

Carpineto, C. and Romano, G. (2004). *Concept Data Analysis: Theory and Applications*, John Wiley, Chichester.

Divya, R., Aswani Kumar, Ch., Saijanani, S., and Priyadharshini, M. (2011). Deceiving communication links on an organization email corpus, *Malaysian Journal of Computer Science* **24**(1): 17–33.

Elloumi, S., Jaam, J., Hasnah, A., Jaoua, A., and Nafkha, I. (2004). A multi-level conceptual data reduction approach based on the Lukasiewicz implication, *Information Sciences* **163**(4): 253-262.

Ganter, B. and Wille, R. (1999). *Formal Concept Analysis: Mathematical Foundations*, Springer, Berlin.

Ghosh, P., Kundu, K. and Sarkar, D. (2010). Fuzzy graph representation of fuzzy concept lattice, *Fuzzy Sets and Systems* **161**(12): 1669–1675.

Horner, V. (2007). *Developing a consumer health informatics decision support system using formal concept analysis*, Masters' thesis, University of Pretoria, Pretoria.

Jamil, S. and Deogun, J.S. (2001). Concept approximations based on rough sets and similarity measures, *International Journal of Applied Mathematics and Computer Science* **11**(3): 655–674.

Kazienko, P. (2009). Mining indirect association rules for web recommendation, *International Journal of Applied Mathematics and Computer Science* **19**(1):165–186, DOI: 10.2478/v10006-009-0015-5.

Liu, M., Shao, M., Zhang, W., and Wu, W.C. (2007). Reduction method for concept lattices based on rough set theory and its application, *Computers and Mathematics with Applications* **53**(9): 1390–1410.

Pattison, P.E. and Breiger, R.L. (2002). Lattices and dimensional representations: Matrix decompositions and ordering structures, *Social Networks* **24**(4): 423–444.

Poelmans, J., Elzinga, P., Viaene, S., Dedene., G. (2010). Formal concept analysis in knowledge discovery: A Survey *in* R. Goebel (Ed.) *Proceedings of the 18th International Conference on Conceptual Structures*, Springer-Verlag, Berlin, pp. 139–153.

Priss, U. (2006). Formal concept analysis in information science, *Annual Review of Information Science and Technology* **40**(1): 521–543.

Snasel, V., Polovincak, M., Dahwa, H.M. and Horak, Z. (2008). On concept lattices and implication bases from reduced contexts, *Proceedings of the ICCS Supplement, Toulouse, France*, pp. 83–90.

Stumme, G. (2009). Formal concept analysis, *in* S. Staab and R. Studer (Eds.), *Handbook on Ontologies*, Springer-Verlag, Berlin, pp. 177–199.

Stumme, G., Wille, R. and Wille, U. (1998). Conceptual knowledge discovery in databases using formal concept analysis methods, *Proceedings of the 2nd European Symposium on Principles of Data Mining and Knowledge Discovery, Nantes, France*, pp. 450–458.

Valtchev P, Missaoul R, and Godin R. (2004). Formal concept analysis for knowledge discovery and data mining: The new challenges, *Proceedings of the 2nd International Conference on Formal Concept Analysis, Sydney, Australia*, pp. 352–371.

Varmuza, K., Filzmoser, P. and Liebmann, B. (2010). Random projection experiments with chemometric data, *Journal of Chemometrics* **24**(3–4): 209–217.

Venter, F.J., Oosthuizen, G.D. and Ross, J.D. (1997). Knowledge discovery in databases using concept lattices, *Experts Systems with Applications* **13**(4): 259–264.

Wille, R. (2002). Why can concept lattices support knowledge discovery in databases?, *Journal of Experimental and Theoretical Artificial Intelligence* **14**(2–3): 81–92.

Wille, R. (2008). Formal concept analysis as an applied lattice theory, *Proceedings of the 4th International Conference on Concept Lattices and Applications, Tunis, Tunisia*, pp. 42-67.

Wu, W.Z., Leung, Y., and Mi, J.S. (2009). Granular computing and knowledge reduction in formal contexts, *IEEE Transactions on Knowledge and Data Engineering* **21**(10): 1461–1474.

Cherukuri Aswani Kumar is an associate professor at the School of Information Technology and Engineering, VIT University, India. He holds a Ph.D. degree from VIT University, as well as a bachelor' and a master's degree in computer science from Nagarjuna University, India. His current research interests are formal concept analysis, data mining and machine intelligence. He has published 32 refereed research papers so far in various international journals and conferences. He was the principal investigator in a major research project funded by the Department of Science and Technology, Government of India, during the period of 2006–2008. Presently he is the principal investigator in a major research project sponsored by the National Board of Higher Mathematics, Department of Atomic Energy, Government of India. He is associated with various professional bodies including ACM, CSI and ISTE. He is a reviewer and an editorial board member of many international journals and conferences.