Reduction dimension of bags of visual words with FCA

Ngoc Bich Dao, Karell Bertet, Arnaud Revel

Laboratoire L3i, University of La Rochelle, France

Abstract. In image retrieval involving bag of visual words, reduction dimension is a fundamental task of data preprocessing. In recent years, several methods have been proposed for supervised and unsupervised cases. In the supervised case, the problem has been addressed with encouraging results. However, in the unsupervised case, reduction dimension is still an unavoidable challenge. In this article, we propose an application of a logic reduction dimension method which is based on Formal Concept Analysis for image retrieval. This method is the reduction of a closure system without, theoretically, loss of information. In our context, combining our proposed method with bag of visual words is original. Experimental results on five data sets such as COREL, CALTECH256, VOC2005, VOC2012 and MIR flickr are analyzed to show the influence of the data structures and the parameters on the reduction factor.

1 Introduction

Thanks to the generalization of multimedia devices, huge collections of digital images are available today. As far as mining in multimedia documents is concerned, web search engines usually give poor results. Hence, such results are far from expected regarding the semantics of the documents. Content Based Image Retrieval (CBIR)[1] has been investigated in order to give an answer to this problem for decades. The main idea is to build a description based on the image content, and to find similarities between descriptions. Classically, visual features are extracted from images and then compiled into an index or signature to give a dense description of images. To perform the retrieval, a similarity function is computed to compare the index of the query with those of collection. A ranking of the results according to the calculated similarity is proposed to the users. The detection of visual features can be performed by a SIFT detector[2] or a dense grid which both select an important number of interest points (up to several thousands) from the images. Each of these points is then described thanks to a SIFT-like descriptor. However, to limit the dimension of the description space, a vector quantization (usually k-means) is performed in order to cluster similar interest points into "visual words", and to generate a dictionary of "visual words" (usually up to 1000 words). Then, the signature of the image is composed of the set of all the visual words corresponding to each feature point detected into the image (what formed a "bag of visual words"[3]). The comparison between the images then consists in comparing the bags of visual words of each image

© Karell Bertet, Sebastian Rudolph (Eds.): CLA 2014, pp. 219–231,
ISBN 978–80–8152–159–1, Institute of Computer Science, Pavol Jozef Šafárik University in Košice, 2014.

in a dataset. The processing cost introduced by these techniques makes them difficult to use with large amounts of images such as a query on the Internet.

On the other hand, supervised data is labeled (the data has ground truth) and classification methods are required to deal with the categorization problem. Data in the case unsupervised is unlabeled, hence clustering methods are used to gather the similar observations in the same cluster. There are many applications for classification and clustering on many domains of computer science such as bioinformatics, numerical analysis, machine learning, data mining, pattern recognition, etc., where data may contain a grand set of features, means the description of the data is high dimension, and therefore it need to be reduced. However, reduction them while preserving the quality of the data is still challenging.

To be able to manage high dimensional description spaces, reduction techniques have been proposed. These techniques are much used as a data preprocessing step in machine learning and pattern recognition. This step can usually increase the accuracy of the results in the next steps such as classification or clustering while the computational cost and time cost of the former step may be significantly decreased. Regarding statistics and machine learning literature, we distinguish two main strategies: feature extraction and feature selection. These methods can be used for supervised case or unsupervised case. The main idea of feature transformation consists in transforming the given set of features into a new one. In case that the size of the new feature set is greater than the original feature set, we called it the feature generation. And when new feature set size is smaller than the original feature set, feature extraction is mentioned. Feature selection methods propose a manipulation of data to select features from the original set. This approach is interesting in some domains when they prefer the existing features in order to maintain their physical properties.

In this article, we propose a logic and unsupervised feature reduction method issued from FCA to address the visual word reduction problem in a CBIR system. In FCA, data are organised into a "context" by a set of observations (called "objects", "samples" or "experimental units" in other fields) and a set of features (also known as "attributes", "parameters", or "variables" in computer science, machine learning and statistic communities) that are associated with each observation.

Context reduction is a simple and polynomial treatment in FCA classically applied on the whole context, thus both reducing observations and features. This treatment is based on a nice result establishing that the concept lattice of the context can be reduced to a minimal one while preserving its graph structure by deleting some redundant observations and features. For example, when two attributes are shared by the same objects, then they belong to the same concepts of the concept lattice, thus they are redundant and one of these two objects can be deleted while preserving the concept lattice structure. In our case, we focus on feature reduction of a context. Our algorithm accepts as input the closure operator of the context on attributes set, and returns the redundant attributes. Thus, this algorithm extends the classical attributes reduction of a context to the more general case of data described by a closure operator. Moreover, we propose a new application in image analysis for features reduction of visual words.

This paper is organized as follows: In order to introduce our approach, we recall some definitions of formal concepts in the section 2.1. Section 2.2 shows details our proposed method. Section 3 shows some experimental results with real data. Finally, section 4 ends this paper with a conclusion and perspectives.

2 The proposed features selection method

The feature reduction algorithm we propose is a logic and unsupervised method stemming from FCA where a concept lattice, defining from a binary table, represents the description of all object-attribute combinations. When the concept lattice structure is preserved after the deletion of some attributes and objects, then these attributes are "redundant" for the lattice structure and can be deleted from the initial data without affecting the structure of object-attributes combinations. Therefore, from a theoretical point of view, the description of data is equivalently represented by a concept lattice where "redundant" attributes and objects are deleted.

The reduction is a simple and polynomial treatment in FCA, classically decomposed into two steps: attribute and object reduction. In this article, we focus on attributes/features reduction, thus on the detection of redundant attributes for the concept lattice structure reduced to attributes. A nice result establishes that each subset of a concept (A,B) is a closure defined on the objects and attributes set, and the concept lattice reduced to the attributes/objects is denoted a closure lattice.

In the first subsection, we introduce the notions of closure lattice according to a closure operator, reduced closure lattice and redundant attributes. In the second section, we presents the reduction algorithm aiming at removing redundant attributes, with a closure operator as input. This algorithm is thus a generic algorithm that can be applied either on attributes or on objects of a binary table, but also on any closure system.

2.1 Reduced lattice

In FCA, the relationship between a set of attributes I and a set of objects O are described by a formal context $(O, I, (\alpha, \beta))$ where $\alpha(A)$ the set of attributes sharing by a subset A of objects, and $\beta(B)$ the set of objects sharing a subset B of attributes. One can derive two closure systems from a context. The first one is defined on the set of attributes I, with $\beta \circ \alpha$ as closure operator. The second one is defined on the set of objects O with $\alpha \circ \beta$ as closure operator[18]. A *closure system* (φ, S) is defined by a *closure operator* φ on a set S, i.e. a map on $\mathcal{P}(S)$ satisfying the three following properties: φ is *isotone*, *extensive* and *idempotent*. A subset $X \subseteq S$ is called *closed* if $\varphi(X) = X$ (see Table 2). The set system $\mathcal{F}$ of all closed subsets, fitted out with the inclusion relation $\subseteq$, forms a lattice usually called the *closure lattice* (see Fig. 1a). See the survey of Caspard and Monjardet[19] for more details about closure systems. There are infinitely set systems whose closure lattice are isomorphic. A *reduced closure lattice* is a closure lattice defined on a set S of the smallest size among all isomorphic closure lattices. A nice result[20,18] establishes that a closure system is reduced when, for each $x \in S$, the closure $\varphi(x)$ is a join irreducible (Equation 1).

$$\forall x \in S, \forall Y \subseteq S \text{ so that } x \notin Y, \text{ then } \varphi(x) \neq \varphi(Y) \tag{1}$$

Therefore, a non-reduced closure system contains reducible elements - elements which do not satisfy Equation 1 - each reducible element $x \in S$ is then equivalent to a set $E_x \subseteq S$ of equivalent elements with $x \notin E_x$ and $\varphi(x) = \varphi(E_x)$. Reducible elements can be removed without affecting the structure of the closure lattice. The reduction of a closure system consists then in removing or replacing each reducible element $x \in S$ by its equivalent set E_x.

2.2 Proposed reduction algorithm

The algorithm we propose is a generic reduction algorithm since it only needs a closure operator as input. Thus it can be applied with the same complexity on any closure system, and in particular on a context by considering the attributes - using $\beta \circ \alpha$ as closure operator.

	a	b	c	d	e	f	g	h
1		×					×	
2	×		×		×		×	
3				×	×	×	×	×
4	×			×	×	×	×	×
5	×				×	×	×	×
6		×				×	×	
7	×					×	×	
8	×				×		×	
9	×	×	×	×	×	×	×	×

(a) The context

	a	b	c	d	e	f
1		×				
2	×		×		×	
3				×	×	×
4	×			×	×	×
5	×				×	×
6		×				×
7	×					×
8	×				×	
9	×	×	×	×	×	×

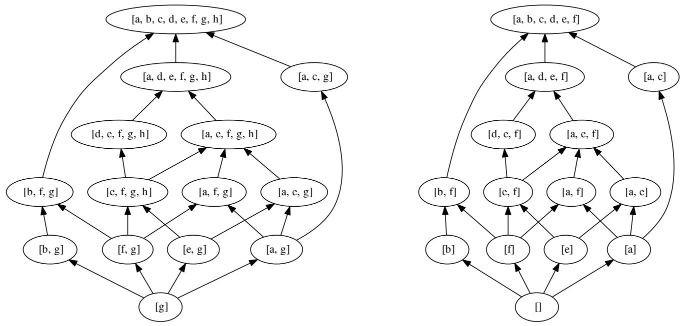
(b) The attribute-reduced context

Table 1: The example of context

x	a	b	c	d	e	f	g	h
$\varphi(x)$	a,g	b,g	a,c,g	d,e,f,g	e,g	f,g	g	e,f,g,h

Table 2: Attributes $x \in S$ and their closure $\varphi(x)$ for the context in Table 1a

A direct application of the definition (see Eq. 1) would imply an exponential cost by checking if any subset $Y \subset S$ is equivalent to each $x \in S$. We use the precedence relation (precedence graph) for a polynomial reduction. The precedence graph is defined on the set S, with an edge between two elements $x, y \in S$



(a) The closure lattice of context in Table 1a

(b) The reduced closure lattice of context in Table 1b

Fig. 1: The example of closure lattices

when $\varphi(x) \subseteq \varphi(y)$. This graph is clearly acyclic for a reduced closure system. We propose a generic algorithm in 3 steps:

Step 1: Standardization. Check if there exists $x, y \in S$ such that $\varphi(x) = \varphi(y)$. When $\varphi(x) = \varphi(y)$, then x and y belong to the same strongly connected components of the graph. Each strongly connected components $X \subseteq S$ include the elements x_i, x_j so that $\varphi(x_i) = \varphi(x_j), \forall x_i \neq x_j \in X$. Thus, we can delete all elements except one representative element $x \in X$ of the component. The obtained precedence graph is then an acyclic graph.

Step 2: Clarification. Check if there exists $x \in S$ such that $\varphi(x) = \varphi(\emptyset)$. When such an x exists, then $\varphi(x)$ is included into $\varphi(y)$ for any $y \in S$, thus x is the only source of the precedence graph. The clarification test has only to be performed for graphs with one source.

Step 3: Reduction. Check, for any $x \in S$, if there exists a set $E_x \subset S$ such that $x \notin E_x$ and $\varphi(x) = \varphi(E_x)$. One can observe that an attribute x with only one immediate predecessor y is not reducible, because it would be equivalent to y, and thus belong to the same strongly connected component already removed in the previous step. If there exists $E_x \subset S$ such that $\varphi(x) = \varphi(E_x)$, then elements of E_x are clearly predecessors of x in the precedence graph since, for $\forall y \in E_x$, $\varphi(x) = \cap\varphi(y)$. Moreover, this test can be reduced to maximal predecessors of x. Therefore, this treatment has only to be performed for elements with more than one immediate predecessors, and the equality has to be checked with the set of immediate predecessors of x.

This algorithm takes into account a closure operator φ on a set S as input. The output of the alforithm is the reducible element set $X \subset S$ and the equivalent elements set E_x for each $x \in X$.

Alg. 1 reduces a closure system in $O(|S|.c_\varphi + |S|^2 \log |S|)$ where c_φ is the cost of a closure generation and —S— is the number of nodes. Indeed, the precedence graph can be initialized in $O(|S|c_\varphi + |S|^2 log|S|)$ by computing the closures in $O(|S|c_\varphi)$, and then comparing two closures in $O(|S|^2 log|S|)$. Then, the SCCs can be computed using Kosaraju's algorithm by two passes of depth first search, thus a complexity in $O(|S| + |A|) \leq O(|S|^2)$, with $|A|$ nb of edges in the graph. Standardization and clarification are clearly in $O(|S|)$ by a simple pass into the graph. Finaly, reduction considers the immediate predecessors of each $x \in S$ in $O(|S|^2)$, and then computes and compare two closures in $O(|S|c_\varphi + |S|^2 log|S|)$. Therefore, Alg. 1 computes the attribute reduced context in $O(|I|^2|O| + |I|^2 log|I|)$. since a closure can be obtained in $O(|I|.|O|)$.

```
Input: a closure operator φ on a set S
Output: the reducible elements set X ⊂ S, and the equivalent elements set E_x
        for each x ∈ X
init a set Res with ∅;
init a graph G with S as set of node;
\\ Precedence graph;
foreach (x, y) ∈ S × S do
    if φ(x) ⊆ φ(y) then
        add the edge (x, y) in G;
    end
end
compute the set CFC of the strongly connected components of G;
let source be the sources of the graph G;
\\ Step (1): Standardization;
foreach C ∈ CFC do
    choose y ∈ C;
    foreach x ∈ C such that x ≠ y do
        add x in Res with E_x = {y}; delete x from the graph G;
    end
end
\\ Step (2): Clarification;
if |source| = 1 and φ(source) = φ(∅) then
    add source in Res with E_source = ∅; delete source from G;
end
\\ Step (3): Reduction;
foreach x ∈ G do
    let P the set of immediate predecessors x in the graph G;
    if |P| ≠ 1 and φ(x) = φ(P) then
        add x in Res with E_x = P; delete x from the graph G;
    end
end
return Res, (E_x)_{x∈Res};
```

Algorithm 1: Reduction of a closure system

3 Experimentation

3.1 Datasets

In our experiments, we compare the performance of the method we propose on different image data sets. Each image in a data set is described by a vector composed of the occurrence frequencies of its visual words, where a set of visual words is defined for each data set. Table 3 describes the different data sets we used in our experiments, and the methods applied to generate the whole bag of visual words.

Database	Images nb	Features nb	Detector	Descriptor	Dictionary of visual words
VOC2012[21]	17124	4096	Harris-Laplace	CMI (Colour Moment Invariants)[22]	Random selection of all key points
MIR flickr[23]	24991	4096	Harris-Laplace	CMI[1]	Random selection of all key points
COREL[24]	4998	500	SIFT	SIFT[2]	K-means[25] (OpenCV)
CALTECH 256[26]	30607	500	SIFT	SIFT[2]	K-means (OpenCV)
Dataset 1 (VOC2005)[27]	1354	262	Harris-Laplace and Laplacian[3]	SIFT	K-means (OpenCV)

Table 3: Description of used datasets

3.2 Experimental protocol

As mentioned earlier, the algorithm we propose requires binary values indicating for each object whether it possesses a given attribute or not. Since each image is described by a visual word occurence frequency vector, its values can vary from 0 to a max value depending on the image size and the quantity of visual words in the image. For instance, if an image is black painted, there is only one visual word "black" for the whole image with a big frequency, and the vector

[1] http://koen.me/research/colordescriptors/
[2] http://www.robots.ox.ac.uk/ vgg/research/affine/#software
[3] http://lear.inrialpes.fr/people/dorko/downloads.html

will be sparse. Conversely, an image with a patchwork of colors is described by a frequency vector mainly composed of low but not zero values. To be able to compare several images, it is thus necessary to normalize their frequency vector before binarization.

Normalization As mentioned before, the visual word occurrence frequency can be very important in some images, and insignificant in others. In order to compare the visual words, several strategies can be adopted.

First of all, it is necessary to find out a "max" value in the data set and then divide the visual word frequency by this max value to transform the values in a range 0 to 1. Two manners to define the max value have been considered into this article.

Normalization by line (image) With this type of normalization, a max value is computed for each image as being the maximum frequency value of the corresponding image. The interpretation of this normalization is that we consider as significant the ratio between the different attributes of a given image. This kind of normalization does not depend on the database size and on the image size. However, the normalized values do not account for the ratio measurement of the same attribute between the images in the database.

Normalization by column (feature) Normalization by column finds out the maximum values of the frequency for each attribute in the database. With this approach, the correspondence between the images in the database is taken into account. The drawback is that each time a new image is inserted into the database, the normalized values must be recomputed. Besides, the image size must also be taken into account. Table 4 gives an illustrated example.

	f_1	f_2	f_3	f_4
img_1	1	0	50	5
img_2	10	9	1	8
img_3	0	0	0	99

(a) Initial data

	f_1	f_2	f_3	f_4
img_1	0.02	0	1	0.1
img_2	1	0.9	0.1	0.8
img_3	0	0	0	1

(b) After normalization by line

	f_1	f_2	f_3	f_4
img_1	0.1	0	1	0.05
img_2	1	1	0.02	0.08
img_3	0	0	0	1

(c) After normalization by column

Table 4: Illustration for normalization types

Binarization After the normalization, we simply binarize the normalized values by comparing these values with a threshold varying from 0 to 0.9. At the highest threshold one, in the *normalization by line* case, it is possible that most of the attributes in an image should be below the threshold. To avoid removing all the visual words from an image, the highest threshold has been assigned to 0.9.

Reduction The next phase in the algorithm is to apply our reduction method which is itself composed of three steps (clarification, standardisation, reduction). Indeed, before applying the proposed method to bag of visual words, we must remove all the visual words that appear (resp. do not appear) in each (resp. any) image. This step corresponds to the *clarification*. The *standardization* step reduces the feature that the vector of images of a given feature equivalent to the vector of images of another feature. At last, in the *reduction* step, all the features which are the combination of other features are removed.

3.3 Results

In this section, we detail the results obtained with our reduction method for 5 data sets, described in section 2.2. To analyze the behavior of our method, and the contribution of each step of the algorithm, we introduce the ratio of removed features for each step of the reduction algorithm as follows:

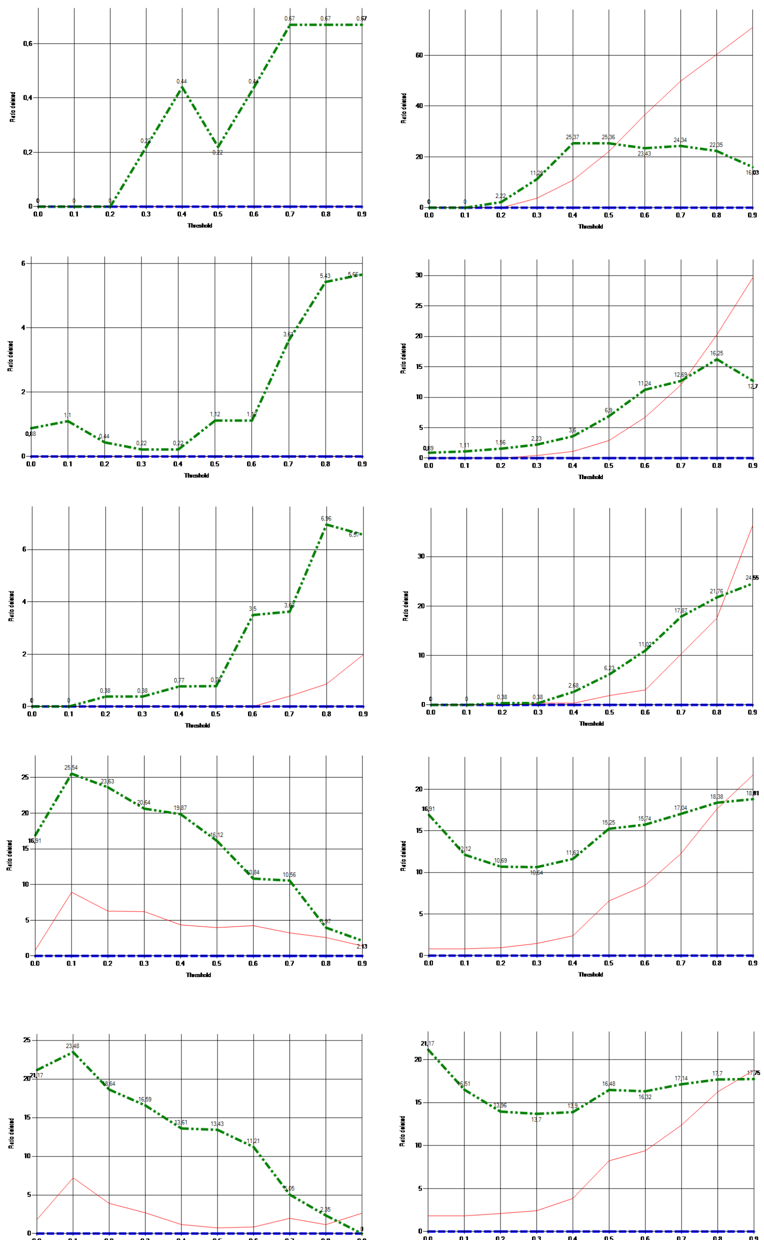
$$\Delta_1 = \frac{a}{N_{att}},\ \Delta_2 = \frac{b}{N_{att}-a},\ \Delta_3 = \frac{c}{N_{att}-a-b}$$

Where a (resp. b and c) is the number of removed attributes in the *standardization* (resp. *clarification* and *reduction*) step; N_{att} is the attribute number in total. Figure 2 shows the evolution of Δ_1, Δ_2, Δ_3 with regard to the threshold level, for both normalization types: line and column.

The maximum ratio of removed attributes of the data sets (CALTECH, COREL, VOC2005, MIRflickr, VOC2012) are approximately equal to 0.67%, 2.6%, 22.5%, 95%, 96% respectively. The impact of the reduction is more interesting in the last three datasets. This phenomenon can be explained by the bag of visual words generation since the two data sets MIR flickr and VOC2012 are composed of randomly selected visual words stemming from the keypoints set. Conversely, the data sets CALTECH, COREL and VOC2005, are composed of bags of visual words defined by the SIFT detector and descriptor, and by a K-means clustering. Thus, the randomly selected visual words are less consistent.

We can also observe that the percentage of removed attributes increases while the binarization threshold increases. With an increasing threshold, only the most frequent words are kept, thus more attributes are potentially equivalent and removed.

At last, there is no attribute reduction in the step 1 (Δ_1 value) with a normalization by column because this kind of normalization can not generate empty columns. Morover, a normalization by line keeps the most frequent attributes in each image whereas a normalization by column keeps the most frequent images for each attribute. To summarize, the number of removed attributes depends both on the visual words generation, on the chosen threshold of binarization and on the normalization process (by line or column). However, care should be taken, that the greater the binarization threshold is, the smaller the number of images remaining. Except in the case normalization by line.



CALTECH

COREL

*VOC*2005

MIRflickr

*VOC*2012

(a) Normalization by line (b) Normalization by column

Fig. 2: The ratio of removed attributes according to the initial attributes corresponding to three cases of proposed method where red line is Δ_1, blue dash is Δ_2 and green dash dot dot is Δ_3.

4 Conclusion and perspective

In this article, we present a logic feature selection method of bags of visual words. This method, stemming from Formal Concept Analysis, is a closure system reduction without, theoretically, loss of information. That means that the data description lattice is preserved by the reduction treatment. In our context, combining our proposed method with a bag of visuals words is original. The experimentations show that the number of deleted features can be interesting, depending on the data set and the binarization treatment. Moreover, it is possible to perform both an object and an attribute reduction.

A finer analysis should be obtained in the supervised case, by comparing classification performance before and after reduction. Moroever, the number of potentially deleted objects could also be usefull to autmatically define a good binarization thresold in the supervised case: while suppression of objects belonging to the same class is to promote, we must avoid removing objects of different classes. Objects reduction can easily be performed by applying our reduction algorithm on the objects set.

At last, we plan to study the number of deleted attributes and deleted objects (of the same class / of different class) to evaluate the complexity of a data set, and the quality of its visuals words.

Acknowledgment: We would like to thank Thierry URRUTY, Nhu Van NGUYEN and Dounia AWAD who extracted the bag of visual words we used in this paper.

References

1. Smeulders, A., Worring, M., Santini, S., Gupta, A., Jain, R.: Content-based image retrieval at the end of the early years. IEEE Transactions on Pattern Analysis and Machine Intelligence **22** (2000) 1349–1380
2. Lowe, D.G.: Object recognition from local scale-invariant features. In: Proceedings of the International Conference on Computer Vision, Kerkyra (1999) 1150–1157
3. Bosch, A., Zisserman, A., Munoz, X.: Scene Classification Via pLSA. In Leonardis, A., Bischof, H., Pinz, A., eds.: 9th European Conference on Computer Vision. Volume 3954 of Lecture Notes in Computer Science., Graz, Austria, Springer Berlin Heidelberg (2006) 517–530
4. Tufféry, S.: Data mining et statistique décisionnelle: L'intelligence des données. Technip edn. Volume 2010. (2010)
5. Belohlavek, R., Kruse, R., Vychodil, V.: Discovery of optimal factors in binary data via a novel method of matrix decomposition. Journal of Computer and System Sciences **76** (2010) 3–20
6. Fisher, R.A.: The use of multiple measurements in taxonomic problems. The Annals of Eugenics **7** (1936) 179–188
7. Hotelling, H.: Analysis of a complex of statistical variables into principal components. Journal of Educational Psychology **24** (1933) 417–441
8. Yu, L., Liu, H.: Feature selection for high-dimensional data: A fast correlation-based filter solution. In: Proceedings of the Twentieth International Conference on Machine Learning (ICML-2003), Washington DC (2003) 856–863

9. Hall, M.A.: Correlation-based feature subset selection for machine learning. Doctor of philosophy, University of Waikato, Hamilton, NewZealand (1999)
10. Battiti, R.: Using mutual information for selecting features in supervised neural net learning. IEEE transactions on neural networks / a publication of the IEEE Neural Networks Council **5** (1994) 537–550
11. Rakotomalala, R., Lallich, S.: Construction d'arbres de decision par optimisation. Revue Extraction des Connaissances et Apprentissage **16** (2002) 685–703
12. Kononenko, I.: Estimating attributes: Analysis and extensions of RELIEF. In Bergadano, F., Raedt, L., eds.: Machine Learning: ECML-94. Volume 784 of Lecture Notes in Computer Science. Springer Berlin Heidelberg, Berlin, Heidelberg (1994) 171–182
13. He, X., Cai, D., Niyogi, P.: Laplacian score for feature selection. In: Neural Information Processing Systems Foundation, MIT Press (2005)
14. Devaney, M., Ram, A.: Efcient Feature Selection in Conceptual Clustering. In: Machine Learning: Proceedings of the Fourteenth International Conference, Nashville, TN (1997)
15. Dy, J.G., Brodley, C.E.: Feature Selection for Unsupervised Learning. Journal of Machine Learning Research **5** (2004) 845–889
16. Wolf, L., Shashua, A.: Feature Selection for Unsupervised and Supervised Inference: The Emergence of Sparsity in a Weight-Based Approach. The Journal of Machine Learning Research **6** (2005) 1855–1887
17. Elghazel, H., Aussem, A.: Unsupervised feature selection with ensemble learning. Machine Learning (2013)
18. Barbut, M., Monjardet, B.: Ordre et classification: algèbre et combinatoire. Hachette, Paris (1970)
19. Caspard, N., Monjardet, B.: The lattices of closure systems, closure operators, and implicational systems on a finite set: a survey. Discrete Applied Mathematics **127** (2003) 241–269
20. Birkhoff, G.: Lattice Theory. 1st edn. American Mathematical Society (1940)
21. Everingham, M., Van Gool, L., Williams, C.K.I., Winn, J., Zisserman, A.: The PASCAL Visual Object Classes (VOC) Challenge (2012)
22. Mindru, F., Tuytelaars, T., Gool, L.V., Moons, T.: Moment invariants for recognition under changing viewpoint and illumination. Computer Vision and Image Understanding **94** (2004) 3–27
23. Huiskes, M.J., Lew, M.S.: The MIR flickr retrieval evaluation. In: Proceeding of the 1st ACM international conference on Multimedia information retrieval - MIR '08, New York, USA, ACM Press (2008) 39–43
24. Carneiro, G., Chan, A.B., Moreno, P.J., Vasconcelos, N.: Supervised Learning of Semantic Classes for Image Annotation and Retrieval. IEEE Transactions on Pattern Analysis and Machine Intelligence **29** (2007) 394–410
25. Macqueen, J.B.: Some Methods for classification and Analysis of Multivariate Observations. In: Proceedings of 5th Berkeley Symposium on Mathematical Statistics and Probability. (1967) 281–297
26. Griffin, G., Holub, A.D., Perona, P.: Caltech-256 Object Category Dataset. Technical report (2007)
27. Everingham, M., Zisserman, A., Williams, C.K.I., Van Gool, L., Al., A.: The 2005 PASCAL Visual Object Classes Challenge. In: First PASCAL Machine Learning Challenges Workshop, MLCW 2005. Volume 3944 of Lecture Notes in Computer Science., Berlin, Heidelberg, Springer Berlin Heidelberg (2005) 117–176