

How to Use the Fractal Dimension to Find Correlations between Attributes

Elaine Parros Machado de Sousa, Caetano Traina Jr., Agma J. M. Traina
University of Sao Paulo at Sao Carlos - Department of Computer Science
Avenida do Trabalhador Saocarlene, 400
13560-970 Sao Carlos, SP - Brazil
Phone: +55-16-273-9693
[parros| caetano| agma]@icmc.sc.usp.br

Christos Faloutsos
Carnegie Mellon University - Department of Computer Science
5000 Forbes Avenue
Pittsburgh, PA 15213-3891
Phone: +1 -412-268-1457
christos@cs.cmu.edu

Abstract

One of the challenges when dealing with multimedia information, which usually is massive and composed of multidimensional data, is how to index, cluster and retain only the relevant information among the items of a large database. In order to reduce the information to the meaningful data, techniques such as attribute selection have been used. In this paper we present a fast, scalable algorithm to quickly select the most important attributes (dimensions) for a given set of n -dimensional vectors, determining what attributes are correlated to the others and how to group them. The algorithm takes advantage of the ‘fractal’ dimension of a data set as a good approximation of its intrinsic dimension and, based on it, indicates what attributes are the most important to keep the meaning of the data. We applied our method on real and synthetic data sets, producing fast and promising results.

Keywords: Attribute selection, dimensionality reduction, intrinsic dimensionality, fractal.

1 . Introduction and Motivation

The volume of information gathered by digital systems has grown not only in the amount of data items but also in the number and complexity of attributes. This happens mostly due to the multimedia nature of the data, and the recent need for storing and retrieving images, video, time series, DNA sequences, among others. When managing large volume of data, a question which frequently arises is: “what part of this data is really relevant to be kept?”. Considering that database management systems stores the information in tables, it is known that usually the relations of the database have many attributes which are correlated between themselves. However, knowing what attributes are the most meaningful and how they are correlated with the other attributes in the relation has been sought by the database and machine learning communities during the last decade.

Attribute selection is a classic goal, which is frequently used to overcome the “dimensionality curse” [13] [10]. A carefully chosen subset of attributes improves the performance and efficacy of a variety of algorithms, including indexing, matching, and

retrieval. This is particularly true with redundant data, as many data sets can largely be well-approximated using fewer dimensions. Moreover, attribute selection can also be seen as a way to compress data, since only the attributes which maintain the essential characteristics of the data are kept [7].

Let $A = \{a_1, a_2, \dots, a_n\}$ denote a relation and a_i its attributes. So, attribute selection algorithms aim to find a subset of attributes $B \subseteq A$ such that B can tightly describe A . Accordingly, there is a set of mappings $M_k(B)$ that can closely approximate the values of every attribute of A which are not in B , that is, $M_k(B) \approx a_k, a_k \in (A - B)$. However, selecting the attribute set that better describes a data set is only an introductory step when gathering knowledge about it. In fact, it is possible to exist a subset of attributes $C_k \subseteq B$ such that $M_k(C_k) \approx a_k, a_k \in (A - B)$. In other words, groups of attributes in B may exist that are correlated with some other attributes of A but that are independent from the other attributes.

In this paper we introduce a novel technique, based on the fractal dimension of the data set, that can discover dependency groups among the most significant attributes to characterize a data set. We also present a fast, scalable algorithm to quickly identify the attributes that determine the dependent attributes.

The paper is organized as follows. Section 2 discusses briefly the previous work on this subject. Section 3 gives an overview of the fundamental concepts concerning the technique we propose in Section 4. Section 5 presents experimental results and Section 6 summarizes our main contributions.

2 . Related work

Several selection methods have been presented in the literature, including genetic algorithms; sequential feature selection algorithms such as forwards, backwards and bi-directional sequential searches; and feature weighting [1] [14] [21]. A survey of attribute selection using machine learning techniques is presented in [3].

As well-known dimensionality reduction techniques one can mention the principal component analysis (PCA) and the singular value decomposition (SVD). These techniques either map the data sets to other spaces or generate a set of additional axes [5]. However, this is not quite attribute selection.

A common research challenge in attribute selection methods is the exponential growth of computing time required [3]. Indeed, the induction methods proposed so far had super-linear and even exponential computational complexity [11], as is the case with nearest neighbors, learning decision trees [9], and Bayesian Networks [16] [12]. Notice that these approaches are highly sensitive to both the number of irrelevant or redundant features presented in the data set and to the size of the data set, avoiding the use of samples [11].

Fractal dimension has been a useful tool to the analysis of spatial access methods [2] [8], indexing [4] [10], join selectivity estimation [6], and analysis of metric trees [17] [18].

3 . Concepts

Database management systems store data in tables where the columns represent the features and the rows represent each data item. Here, we refer to these tables as data sets, to the attributes as features, and to the data elements (or objects) as points in the feature space. Therefore, each data element can be considered as a point in an E -dimensional space, where E is the number of attributes in the data set. Complex data, such as features extracted from multimedia data, are well-known examples of high-dimensional data used in content-based retrieval systems. For these data sets it is difficult to choose the set of attributes that can be assigned as the data set keys. In consequence, if one is interested in creating an index structure on the data the whole set of attributes has to be considered. This leads to the previously mentioned dimensionality curse.

3.1 - ‘Embedding’ and ‘intrinsic dimensionality’

The technique we propose in this work has two objectives: (1) to find a subset of the attributes that, besides concisely representing the data, can be used when creating indexes or applying data mining techniques over the data without compromising the results; (2) to establish which attributes are correlated between themselves.

Although the existence of correlations between attributes in the data set is frequent, it is difficult to know which attributes are correlated with, or are functions of, the others. Namely, it is difficult to establish the functional dependency. Thus, our work aims at detecting the correlations between attributes in a data set, spotting the attributes that can be obtained by some function of others, and defining the reduced data set with only the relevant attributes. These considerations lead to the definition of the embedding and intrinsic dimensions.

Definition 1 - The embedding dimension E of a data set is the dimension of its address space. In other words, it is the number of attributes of the data set.

The data set can represent a spatial object that has a dimension lower than the space where it is embedded. For example, a line has an intrinsic dimensionality one, regardless of whether it is in a higher dimensional space.

Definition 2 - The intrinsic dimension of a data set is the dimension of the spatial object represented by the data set, regardless of the space where it is embedded.

Note that if a data set has all of its features independent of the others, its intrinsic dimension is the embedding dimension. However, whenever there is a correlation between two or more features, the intrinsic dimensionality of the data set is reduced

accordingly. For example, each polynomial correlation (linear, quadratic, and so on) reduces the intrinsic dimension by one unit. Other types of correlations can reduce the intrinsic dimension by different amounts, even by fractional amounts, as we show following.

Usually the embedding dimensionality of the data set hides its actual characteristics. In fact, correlations between the features in real data sets are normally not known, and even the existence of correlations is unknown. Through the intrinsic dimension of a data set it is possible to decide how many attributes are actually required to characterize it. Henceforth, we get the intrinsic dimension D of a data set by the Correlation Fractal Dimension of it (D_2).

Definition 3 (Correlation Fractal Dimension): Given a data set that has the self-similarity property in the range of scales $\{r_1, r_2\}$, its Correlation Fractal dimension D_2 for this range is measured as

$$D_2 \equiv \frac{\partial \log \sum_i C_{r,i}^2}{\partial \log r}, \quad r \in [r_1, r_2] \quad (1)$$

where r is the side of the cells imposed over the data set and $C_{r,i}$ is the count of points in each cell i , as described in [6] [19].

As many real data sets are self-similar, or fractal [15], we can use their correlation fractal dimension as their intrinsic dimension D .

If a data set is stored as a relational table, it can be useful to compute the fractal dimension of a subset of its attributes. The following definition, first presented in [19], handles this situation.

Definition 4 - Partial fractal dimension (pD): Given a data set A with E attributes and a subset of these attributes $B \subset A$, the Partial Fractal Dimension $pD(B)$ is obtained through the calculation of the correlation fractal dimension of the data set considering only the attributes from the B subset.

Algorithms to calculate D_2 and any $pD(B)$ were proposed in [19]. Both algorithms are linear on the number of attributes and on the number of elements in the data set.

4 . Proposed Technique

The proposed technique is based on the concepts we define as follows.

4.1 - Definitions

Let $A = \{a_1, a_2, \dots, a_E\}$ denote a data set composed of only numerical, continuous (non-categorical) attributes. Then, we present the following definitions.

Definition 5 (Correlation super-group) - Given the data set $A = \{a_1, a_2, \dots, a_E\}$ composed by E attributes, a Correlation Super-Group $SG \subseteq A$ is a subset of attributes of A such that, for every attribute $a_k \in SG$ there is a mapping $M_k(SG - \{a_k\}) \rightarrow a_k$ and for any given combination of values of $SG - \{a_k\}$ there are few possible values for a_k , that is, the attributes $SG - \{a_k\}$ are correlated to the attribute a_k .

Definition 6 (Correlation group) - Given a data set $A = \{a_1, a_2, \dots, a_E\}$, a Correlation Group $G \subseteq A$ is a subset of attributes of A such that G is a correlation super-group of A , and there is no attribute $a_j \in G$ such that $G - \{a_j\}$ is a correlation super-group in A .

Experimental conjecture: If there is a function $f_k(G - \{a_k\}) \approx a_k$, $G \subset A$, $a_k \in G$, then the attributes in G are correlated. However,

a correlation is not limited to functions. In fact, any mapping $M_k(\mathcal{G}\{a_k\}) \rightarrow a_k$ that restricts the possible values of the attribute a_k determines a correlation. If the mapping $M_k(\mathcal{G}\{a_k\}) \rightarrow a_k$ is a function, that is, if there is only one value of a_k for a given combination of values of $\mathcal{G}\{a_k\}$, then $pD(\mathcal{G}) = pD(\mathcal{G}\{a_k\})$. By experimental evaluation, we found that, if $M_k(\mathcal{G}\{a_k\}) \rightarrow a_k$ restricts the subset of values of a_k for a given combination of values of $M_k(\mathcal{G}\{a_k\})$ to a few elements, $pD(\mathcal{G}) - pD(\mathcal{G}\{a_k\})$ is also almost zero, increasing slightly as the number of possible values for a_k rises.

Corollary 1 (derived from the experimental evidence) - For every attribute $a_k \in \mathcal{G}$, $pD(\mathcal{G}) - pD(\mathcal{G}\{a_k\}) \approx 0$.

Definition 7 (Correlation base) - If \mathcal{B} is a correlation group in A , and therefore there is no pair of attribute $a_k, a_j \in \mathcal{B}$ such that $pD(\mathcal{B}) - pD(\mathcal{B}\{a_k, a_j\}) \approx 0$, then for every $a_i \in \mathcal{B}$ the set of attributes $\mathcal{B}\{a_i\}$ is a correlation base of A .

The intuition about a correlation group \mathcal{G} with g attributes is that at least one of its attributes is correlated with every other attribute in the group. However, it is possible that less than $g-1$ attributes are enough to define the correlation. So, a correlation base \mathcal{B} , $\mathcal{B} \subset \mathcal{G}$ with b attributes, determines the minimum number of attributes in \mathcal{G} that is needed to define the correlation. In this way, any b attributes in the correlation group determine a mapping from these attributes to every other $g-b$ attributes in the group.

Consider the data set $A = \{a_1, a_2, \dots, a_n\}$ composed only by numerical non-categorical attributes with a fractal dimension D . If a new attribute a_p is added to this data set, the total fractal dimension will increase at most one unit. On the one hand, the total fractal dimension will increase by one unit if a_p is completely uncorrelated with all the other attributes in the data set. On the other hand, if a_p is correlated with the existing attributes, the fractal dimension will increase by a value of almost zero. In a third case, the total fractal dimension will increase by an amount between zero and one if the new attribute a_p is partially, or 'fractally', correlated with the attributes that already exist in the data set.

In the same way, the pD of only one numerical attribute of a data set can also be a value from zero to at most one unit. For a constant attribute, the pD is zero. In addition, if there is no formation law in the distribution of the values of this attribute, its partial fractal dimension will be one. Conversely, if there is a distribution rule for its values, its pD will be a value between zero and one, reflecting the fractal formation of its distribution.

4.2 - The Attribute Significance Estimator

Based on the previous concepts, we developed an algorithm named *Fractal Dimension Attribute Significance Estimator* (FD-ASE). It uses the approach of attribute forward inclusion to recreate the data set. The fundamental idea is to calculate the partial correlation fractal dimension of data set projections, integrating more and more attributes until its calculated correlation fractal dimension D is achieved. An outline of the FD-ASE algorithm is shown in Figure 1.

The algorithm works as follows: given the data set $A = \{a_1, a_2, \dots, a_n\}$, composed by n attributes identified by an index ranging from 1 to n , the first step is to discard the constant

attributes. To do this, the individual pD of each attribute is computed and every attribute with $pD = 0$ is dropped. For presentation clarity, we consider that no attribute is discarded in step 1.

In the first interaction of the algorithm, step 2 calculates the pD of the partial data set consisting of the attribute a_1 , and progressively calculate the pD of the partial data set resulting from including attributes a_2, a_3 , and so on, until the correlation fractal dimension D of the complete data set is achieved. However, it is possible that, when adding an attribute a_k , the fractal dimension does not increase. This means that the attribute a_k is correlated with at least one of the attributes $a_i \mid i < k$, so attributes a_1 to a_k define a correlation supergroup.

Input: the data set $A = \{a_1, a_2, \dots, a_n\}$ composed by n attributes.
Output: the power set $\{\mathcal{G}_1, \mathcal{G}_2, \dots\}$ of correlation groups in A .

1. Remove every attribute a_i in A whose $pD(a_i) \approx 0$, and set $c=1$.
2. Calculate $pD(a_1), pD(a_1, a_2), \dots$ until $pD(a_1..a_k) \approx pD(a_1..a_{k-1})$.
Set $\mathcal{G}_c = \{a_1, \dots, a_k\}$.
3. For j from 1 to $k-1$ calculate $pD(a_1..a_{j-1}, a_{j+1}..a_{k-1}), pD(a_1..a_{j-1}, a_{j+1}..a_k)$ and $pD(a_1..a_{j-1}, a_{j+1}..a_k, a_j)$.
4. If $pD(a_1..a_{j-1}, a_{j+1}..a_k) \approx pD(a_1..a_{j-1}, a_{j+1}..a_{k-1}) < pD(a_1..a_{j-1}, a_{j+1}..a_k, a_j)$ remove a_j from \mathcal{G}_c .
5. For j from $k+1$ to n calculate $pD(\mathcal{G}_c, a_j)$, and if $pD(\mathcal{G}_c, a_j) \approx pD(\mathcal{G}_c)$ include a_j in \mathcal{G}_c .
6. Reserve \mathcal{G}_c and remove $\mathcal{G}_c - \mathcal{B}$ from A , where \mathcal{B} is the correlation base of \mathcal{G}_c .
7. If there are attributes in A , increase c and repeat from step

Figure 1 - The *Fractal Dimension Attribute Significance Estimator* (FD-ASE) algorithm.

Having found a correlation super-group, it is necessary to identify the correlation group where the attribute a_k pertains, and dropping attributes that are not part of it. We assume, without loss of generality, that there is no other attribute $a_j, j < k$, such that its inclusion in the data set does not change the resulting pD . That is, a_k is the first attribute in the sequence (a_1, a_2, \dots, a_n) such that $pD(a_1..a_k) \approx pD(a_1..a_{k-1})$. By the corollary 1, if the attributes in the sequence $(a_1..a_k)$ constitute a correlation group, there is no attribute in it that can be removed resulting in a different pD . So, in step 3 the algorithm processes the sequence $(a_2..a_k, a_1)$, that is, puts the attribute a_1 as the last element in the sequence, maintaining the other attributes in the same order. In the next step, if $pD(a_2..a_k) \approx pD(a_2..a_{k-1}) < pD(a_2..a_k, a_1)$, then a_1 is not an attribute in the same correlation group with a_k , and the attribute a_1 is removed from the group; otherwise a_1 and a_k are in fact in the same correlation group. Steps 3 and 4 must be repeated for every attribute a_j in $(a_1..a_k)$, computing $pD(a_1..a_{j-1}, a_{j+1}..a_{k-1}), pD(a_1..a_{j-1}, a_{j+1}..a_k)$ and $pD(a_1..a_{j-1}, a_{j+1}..a_k, a_j)$. Whenever an attribute is found not pertaining to the target correlation group, it is eliminated from the sequence in the remaining iterations. Therefore, after at most $3*k$ calculations of the $pD()$ algorithm, the first correlation group of attributes is discovered.

In step 5, the other attributes a_j in $A, j > k$, are tested to determine if they pertain to the same correlation group as a_k . Therefore, every attribute that does not increase the pD of the correlation group already found is included in this group. In step 6, the correlation group \mathcal{G}_1 already found is reserved. Being g the number of attributes in the group \mathcal{G}_1 , only the b attributes that

compose the correlation base of the group is kept for the next execution of the algorithm. That is, g - b attributes are removed from the set of attributes being processed, and the whole algorithm is repeated (step 7), until every correlation groups are discovered.

5. Experimental Results

In order to evaluate the efficacy of our method, we have made several experiments with different synthetic and real data sets. As we cannot discuss all of them here, we selected a few but representative experimental results to demonstrate the main idea of our technique. The results are graphically demonstrated through FD-ASE plots, which are plots in log-log scale of the sum of the number of points $\sum C_{r,i}^2$ versus r , as defined in equation (1). The slope of the line that best fits the plot as each attribute a_k is included corresponds to the pD of the sequence $\{a_1, a_2, \dots, a_k\}$. Note that the sequence plot is presented in the graphs from right to left.

Cube1 - this is a 10,000 points randomly generated data set in a three-dimensional manifold. Its three coordinates are completely uncorrelated to each other.

Figure 2 presents the graph generated by the FD-ASE algorithm on the sequence (a_1, a_2, a_3) . From this plot, it can be seen that each attribute a_i added to the data set increases the pD by one unit, achieving the intrinsic dimension of the whole data set

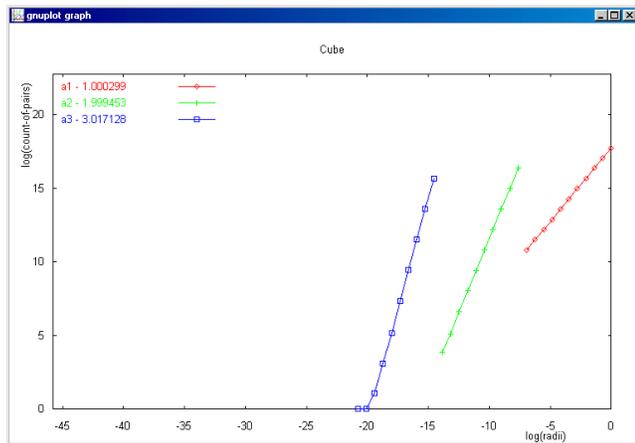


Figure 2 - FD-ASE plot for Cube1.

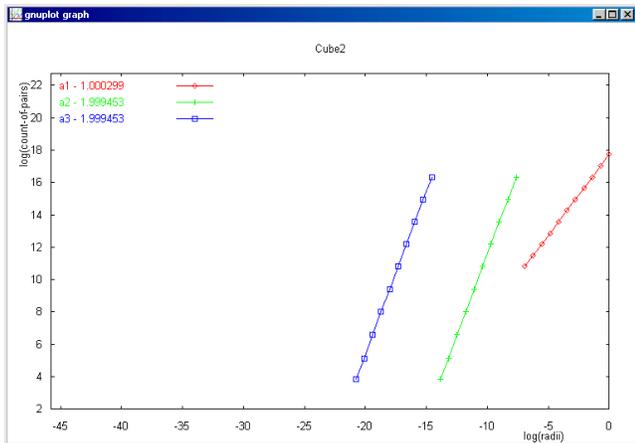


Figure 3(a) - FD-ASE plot for Cube2: sequence (a_1, a_2, a_3) .

when including the attribute a_3 .

Cube2 - this is a three-dimensional data set with 10,000 points and two attributes strongly correlated ($a_1, a_2, a_3=f(a_1)$). Figure 3 presents the graphs of two sequences: (a) (a_1, a_2, a_3) ; (b) (a_1, a_3, a_2) .

Figure 3(a) shows that the pD does not increase when adding attribute a_3 . This means that it is correlated to at least one of the

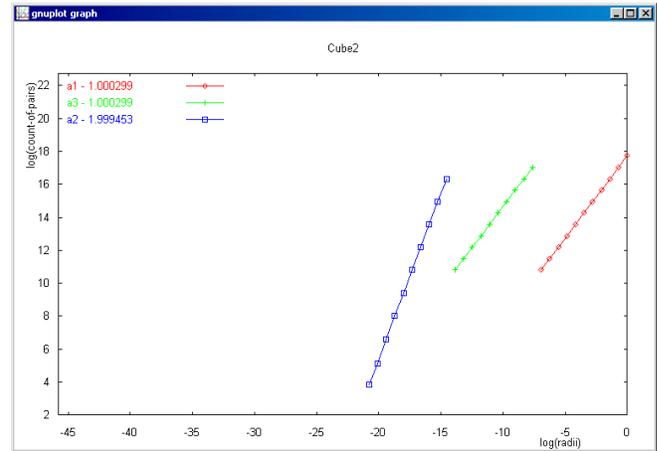


Figure 3(b) - FD-ASE plot for Cube2: sequence (a_1, a_3, a_2) .

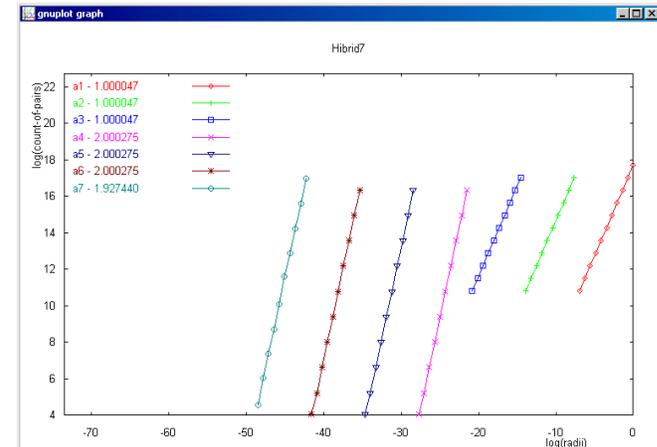


Figure 4(a) - FD-ASE plot for Hibrid7: sequence $(a_1, a_2, a_3, a_4, a_5, a_6, a_7)$.

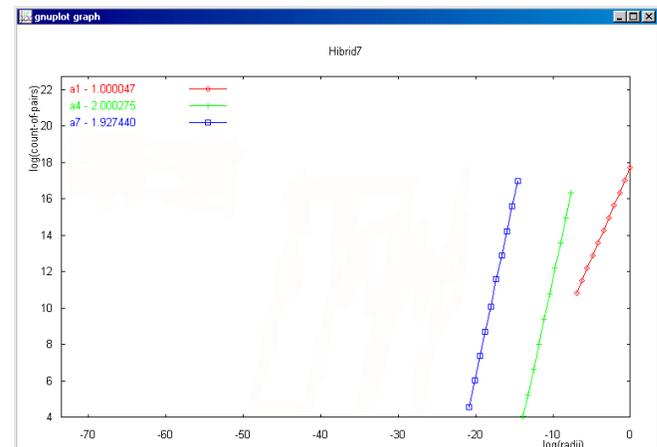


Figure 4(b) - FD-ASE plot for Hibrid7: sequence (a_1, a_4, a_7) .

previous attributes, and therefore the attributes a_1 to a_3 define a correlation super-group. In Figure 3(b), it can be seen that the attribute a_2 , transferred to the end of the sequence, increases the pD by one unit, therefore it is not in the target correlation group. It can also be seen that the increase of the pD when a_3 is added is zero, revealing that attributes a_1 and a_3 are completely correlated.

Hibrid7 - this is a 10,000 seven-dimensional point set, whose attributes are correlated as: $\{a_1, a_2=f(a_1), a_3=f(a_1), a_4, a_5=f(a_4), a_6=f(a_4), a_7=f(a_1, a_4)\}$.

Figure 4 presents graphs generated by the FD-ASE algorithm on two different sequences. Figure 4(a) illustrates the first execution of the algorithm, which results in the first correlation group $G_1 = \{a_1, a_2, a_3\}$. The correlation base of G_1 , $B_1 = \{a_1\}$, is kept in the set of attributes for the next execution, which find $G_2 = \{a_4, a_5, a_6\}$ and $B_2 = \{a_4\}$. At last, the algorithm execution applied to the sequence (a_1, a_4, a_7) determines the third correlation group, as illustrated in Figure 4(b).

Eigenfaces - this is a real data set with 11,900 eigenvalues extracted from human face images. Each face was processed with the eigenfaces method [20], resulting in 16-dimensional vectors.

Although the points of the Eigenfaces data set are embedded in a 16-dimensional space, its intrinsic dimension is 4.25. Figure 5 presents the graph of a sequence processed by the FD-ASE algorithm. From the graph, it can be seen that from the 16 attributes, only 5 of them significantly increase de pD , reflecting the fractal formation of the data set. Executing the FD-ASE algorithm using this data set discovered just one correlation group, what is consistent with the eigenfaces method employed to generate the data set.

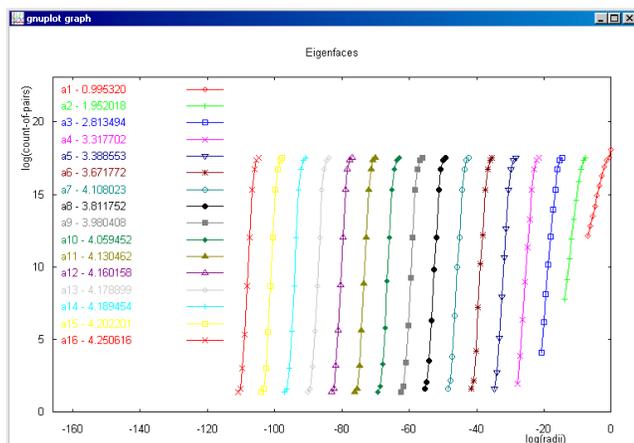


Figure 5 - FD-ASE plot for Eigenfaces.

6. Conclusions

In this paper we presented a novel technique to discover groups of correlated attributes in a data set, using the fractal dimension concept. We also presented an algorithm to find correlation groups that is linear on the number of attributes and on the number of elements in the data set. In this sense, our technique represents an important achievement on finding information about a data set, as it can discover if there is more than one group of attributes that establishes the correlations in the data set, and also identify the attributes that pertain to each group. To the best of our knowledge,

this information cannot be obtained by the existing techniques. Using our proposed method and the presented algorithm, we are now studying a new technique to improve existing methods of dimensionality reduction based on the fractal dimension.

Acknowledges

This research has been partially supported by FAPESP (Fundação de Amparo à Pesquisa do Estado de São Paulo, Brazil - under grant 01/03370-6), and by CNPq (Conselho Nacional de Desenvolvimento Científico e Tecnológico, Brazil - under grants 52.1685/98-6, 52.1267/96-0 and 860.068/00-7).

References

- [1] D. W. Aha and R. L. Bankert, "A Comparative Evaluation of Sequential Feature Selection Algorithms," in Proceedings of the Fifth Intl. Ws on AI and Statistics, Ft. Lauderdale-FL, 1995.
- [2] A. Belussi and C. Faloutsos, "Estimating the Selectivity of Spatial Queries Using the Correlation Fractal Dimension," in VLDB, 1995, pp. 299-310.
- [3] A. Blum, P. Langley, "Selection of Relevant Features and Examples in Machine Learning," *AI*, vol. 97, pp. 245-271, 1997.
- [4] C. Böhm and H.-P. Kriegel, "Dynamically Optimizing High-Dimensional Index Structures," in 7th EDBT, Konstanz, Germany, 2000, pp. 36-50.
- [5] C. Faloutsos, *Searching Multimedia Databases by Content*. Boston, MA: Kluwer Academic Publishers, 1996.
- [6] C. Faloutsos, B. Seeger, A. J. M. Traina, C. Traina, Jr., "Spatial Join Selectivity Using Power Laws," in ACM SIGMOD, Dallas, TX, 2000, pp. 177-188.
- [7] U. M. Fayyad, "Mining Databases: Towards Algorithms for Knowledge Discovery," *Bulletin of Tech. Committee on Data Engineering*, vol. 21, pp. 29-48, 1998.
- [8] I. Kamel and C. Faloutsos, "Hilbert R-tree: An Improved R-tree using Fractals," in 20th VLDB, Chile, 1994, pp. 500-509.
- [9] K. Kira and L. A. Rendell, "A Practical Approach fo Feature Selection," in 9th Intl. Conf. on Machine Learning, Aberdeen Scotland, 1992, pp. 249-256.
- [10] F. Korn, B.-U. Pagel, C. Faloutsos, "On the 'Dimensionality Curse' and the 'Self-Similarity Blessing'," *IEEE TKDE*, vol. 13, pp. 96-111, 2001.
- [11] P. Langley; S. Sage, "Scaling to Domains with many Irrelevant Features," in *Computational learning theory and natural learning systems*, vol. 4, Cambridge, MA: MIT Press, 1997.
- [12] D. Margaritis, C. Faloutsos, S. Thrun, "NetCube: A Scalable Tool for Fast Data Mining and Compression," in VLDB, Roma, Italy, 2001, pp. 311-320.
- [13] B.-U. Pagel, F. Korn, C. Faloutsos, "Deflating the Dimensionality Curse Using Multiple Fractal Dimensions," in 16th ICDE2000, San Diego, CA, 2000, pp. 589-598.
- [14] M. Scherf and W. Brauer, "Feature Selection by Means of a Feature Weighting Approach," Technische Universität München, Munich 1997.
- [15] M. Schroeder, *Fractals, Chaos, Power Laws*, 6 ed. New York: W.H. Freeman and Company, 1991.
- [16] M. Singh and G. M. Provan, "A Comparison of Induction Algorithms for Selective and Non-selective Bayesian Classifiers," in 12th Intl. Conf. on Machine Learning, Lake Tahoe, CA, USA, 1995, pp. 497-505.
- [17] C. Traina, Jr., A. J. M. Traina, C. Faloutsos, "Distance Exponent: a New Concept for Selectivity Estimation in Metric Trees," in IEEE ICDE, San Diego - CA, 2000, pp. 195.
- [18] C. Traina, Jr., A. J. M. Traina, C. Faloutsos, B. Seeger, "Fast Indexing and Visualization of Metric Data sets Using Slim-trees," *IEEE TKDE*, vol. 14, pp. 244-260, 2002.
- [19] C. Traina, Jr., A. J. M. Traina, L. Wu, C. Faloutsos, "Fast feature selection using fractal dimension," in XV Brazilian DB Symposium, João Pessoa - PA - Brazil, 2000, pp. 158-171.
- [20] M. Turk and A. Pentland, "Eigenfaces for Recognition," *Journal of Cognitive Neuroscience*, vol. 3, pp. 71-86, 1991.
- [21] H. Vafaie and K. A. D. Jong, "Robust Feature Selection Algorithms," in Intl. Conf. onTools with AI, Boston, MA, 1993.