

**Semantic Pattern Recognition Based on Linear Algebra and
Latent Semanti Analysis
Amjed Abbas Ahmed**

**Semantic Pattern Recognition Based on Linear Algebra and Latent
Semantic Analysis**

Amjed Abbas Ahmed

Al-Imam Al-Kadhumi College for Islamic Science (Department Of Diyala) Iraq-Baghdad

Received: 5 September 2015

Accepted: 24 March 2016

Abstract

Pattern recognition is a process of identifying vector of correlated/uncorrelated attributes and discriminate it among other patterns. Pattern recognition is synonymous to machine learning, data mining and Knowledge Discovery in Database (KDD). In this research work we investigate decomposing pattern (i.e., attribute vector) space into subspaces in which patterns cluster around basis of the subspaces. This paper introduces a theory which states that in case of having space of vectors and having basis then Signal Value Decomposition (SVD) can perform excellent in discovering thesis basis, hence, in pattern recognition a space can be decomposed to sub-spaces to reach clustering around basis. Results are collected and discussed and it has proven that SVD and its extension Latent Segment Analysis (LSA) can optimize the process of machine learning and showed a great tendency to converge toward cognitive based recognition.

Keywords: pattern recognition, semantic analysis, Singular Value Decomposition (SVD), Latent Semantic Analysis.

**Semantic Pattern Recognition Based on Linear Algebra and
Latent Semanti Analysis
Amjed Abbas Ahmed**

تميز نمط الدلالي بالاعتماد على الجبر الخطي والتحليل الدلالي الكامن

امجد عباس احمد

كلية الأمام الكاظم (عليه السلام) - أقسام ديالى

الخلاصة

تميز الأنماط هو عملية التعرف على الأنماط بتحديد صفات الترابط /عدم الترابط وتمييز ذلك بين الأنماط الأخرى وهو مرادف لتعليم الآلة لاستخراج البيانات واكتشاف المعرفة في قاعدة البيانات. في هذا العمل تم التحقق من تميز الأنماط من خلال تحليل فضاء المدخلات وعلاقته مع الفضاءات الجزيئية باستخدام خوارزمية تفكيك على أساس القيمة المفردة (SVD) وهذه الخوارزمية ممكن أن تؤدي بشكل ممتاز إلى اكتشاف أنماط حول أساس الفضاءات الجزيئية. أن النتائج ومناقشتها أثبتت أن خوارزمية (SVD) والتحليل الدلالي الكامن (LSA) ممكن ان تحسن عملية تعلم الآلة وميل كبير لإدراك التمييز.

كلمات مفتاحيه : تميز الأنماط ، التحليل الدلالي ، تفكيك على أساس القيمة المفردة ، التحليل الدلالي الكامن.

Introduction

Pattern recognition is defined generally as the assignment of a label to a given fixed size input stream of attributes [1], for example the classification process is an outcome of pattern recognition. Pattern recognition has its origin in engineering while machine learning has its origin in computer science, anyway, huge number of applications exploit the methodologies of pattern recognition to accomplish its tasks .

Conventionally, pattern is a vector of N-Dimensions column vector called a feature vector. Feature and attribute are used interchangeably in publishing; this vector is called the vector space which describes a state of a system, a collection of these vectors represent the system space [1,2]. Studies [1,2,3,4,5] have been forwarded to investigate the characteristics of the components composing the vector space, for example the linearity in dependency among vector's components and the correlation in sub-spaces of the problem space. In this paper we investigate the decomposing of the problem space into sub-spaces or clusters based on the outcome of SVD and LSA analysis. The problem domain for this research is the recognition of hand written scripts which are represented by vector of attributes (i.e., attributes are statistical

**Semantic Pattern Recognition Based on Linear Algebra and
Latent Semanti Analysis
Amjed Abbas Ahmed**

calculations that characterize each vector space). In [4] a description for pattern classifiers is presented in which pattern classifiers are usually based on heuristic feature extraction in order to grant simple classifier the ability to perform with high accuracy. The features frequently used in character recognition include the chain code feature, K-L expansion, the Gabor transform and many others. The more discriminative feature extracted from the image the more accurate the performance of the classifier. Clustering techniques also used to increase the performance of pattern recognition due to the reduction of the search space where problem space is categorized into classes where each individual belongs to certain class is close to other vector space in that class; this can be interpreted in different schemes based on the methodology used to label these classes [4,5]. In our approach presented in this paper we will analyze problem space in the semantic space in order to reveal hidden relationships among features (i.e., attributes) of the vector space.

LSA is an analysis tool used to analyze the relation between structures of concepts and the documents containing these concepts [1,5,7]. LSA is not used only with natural language but it can be used with any collection of documents that are composed of structures (i.e., items). LSA use partial information of the image to constitute the methods to achieve global implications. The features of the local information reflected as the text of the meaning of a particular string is still called semantic. The semantic features are not expressed in specific form of intuition, but specific from implicit in the data. The purpose of the LSA used in image analysis is an attempt to extract the high-level semantic concepts from the visual features of the image. And hope it can fill the gap between the low content features and high-level features.

1- Latent Semantic Analysis (LSA)

LSA is a theory and method for extracting and representing the meaning of words. Meaning is estimated using statistical computations applied to a large corpus of text [4]. The corpus embodies a set of mutual constraints that largely determine the semantic similarity of words and sets of words. These constraints can be solved using linear algebra methods, in particular, singular value decomposition [4, 5].

LSA has been shown to reflect human knowledge in a variety of ways. For example, LSA measures correlate highly with humans' scores on standard vocabulary and subject matter tests;

**Semantic Pattern Recognition Based on Linear Algebra and
Latent Semanti Analysis
Amjed Abbas Ahmed**

it mimics human word sorting and category judgments; it simulates word-word and passage-word lexical priming data; and it accurately estimates passage coherence [4, 5].

The core processing in LSA is to decompose A using SVD; SVD has designed to reduce a dataset containing a large number of values to a dataset containing significantly fewer values, but which still contains a large fraction of the variability present in the original data [3, 4, 5].

$$A = U\Sigma V^T \quad \text{---} \quad (1)$$

Where

- 1- $EigenVector(AA^T) \rightarrow Columns(U)$
- 2- $EigenVector(A^T A) \rightarrow Columns(V)$
- 3- $EigenValue(A^T A) \text{ OR } EigenValue(AA^T) \rightarrow \Sigma$

the first structure is the single pattern that represent the most variance in the data, after all, SVD is an orthogonal analysis for dataset, U is composed of eigenvectors of the variance-covariance matrix of the data, where the first eigenvector points to the direction which holds the most variability produced by all other vectors jointly. U is an orthogonal matrix where all its structures are mutually uncorrelated. Eigne values are representing scalar variance of corresponding eigenvectors; this way total variation exhibited by the data is the sum of all eigenvalues and singular values are the square root of the eigenvalues [4, 6,7, 8].

In LSA data is subjected to two-part transformation:

- 1- The word frequency (+1) in each cell is converted to its log.
- 2- The information-theoretic measure, entropy, of each word is computed as $(P \log P)$ over all entries in its row and each cell entry then divided by the row entropy value.

The mentioned two parts transformation is crucial to build the semantic space of the system modeled by the matrix, where, words or features are weighted as an estimate of its importance in the passage [4].

Semantic Pattern Recognition Based on Linear Algebra and Latent Semanti Analysis
Amjed Abbas Ahmed

2- Local Ternary Patterns (LTP)

In [5][6] a detailed description for the LTP (Local Ternary Patterns) is presented, where researchers agreed on using LTP as highly discriminative features for texture classification and as a highly resistivity for lighting effects. Patterns are expressed here in three values fashion where each pixel within the original pattern is represented by the following function:

$$F(p) \begin{cases} 1 & \text{if } p > c+K \\ 0 & \text{if } p > c-K \text{ and } p < c+k \\ -1 & \text{if } p < c-k \end{cases} \quad (2)$$

This approach of representing patterns will results in a better analysis and details as it is clear when subjecting these patterns to Histogram analysis.

Hypothesis to be investigated by this paper is that ‘Semantic attributes vector for the original pattern is a basis vector in a subspace’, which means that all other vectors are spanned by the original pattern. Figure (1) shows the representation of the patterns in the geometrical space and it is shown that vectors can be recognized by interpreting the results of the inner product.

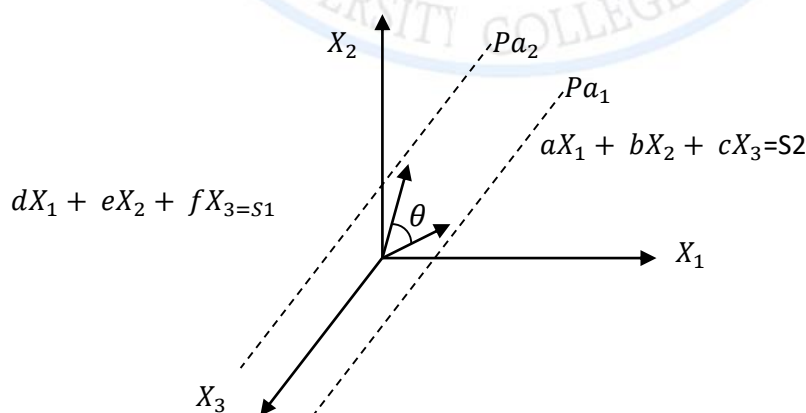
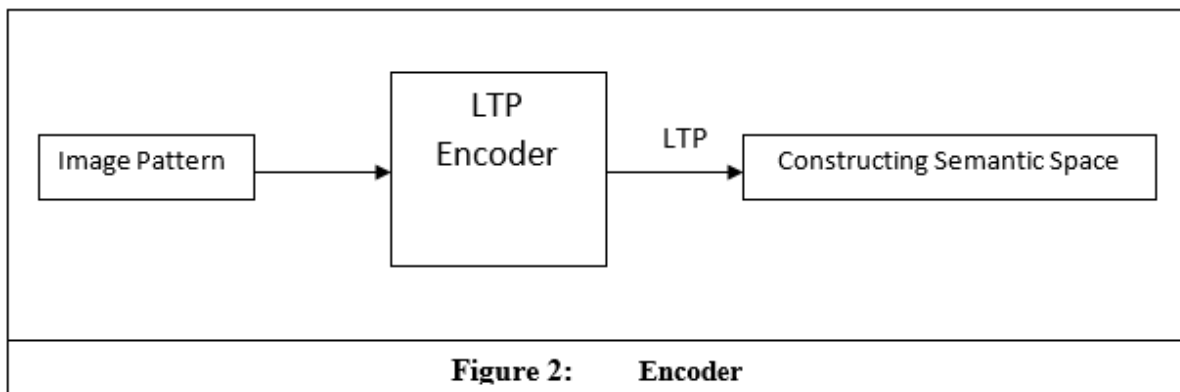


Figure 1: Geometrical Representation for the Proposed Pattern Recognition

**Semantic Pattern Recognition Based on Linear Algebra and
Latent Semanti Analysis
Amjed Abbas Ahmed**

3- The proposed Pattern Recognition Scheme

The proposed pattern recognition system is based on retrieving the latent semantic relationships among attributes that characterize the image (i.e., Mean, Std. deviation, variance, Mode, Median, Max value, Min value and Length).

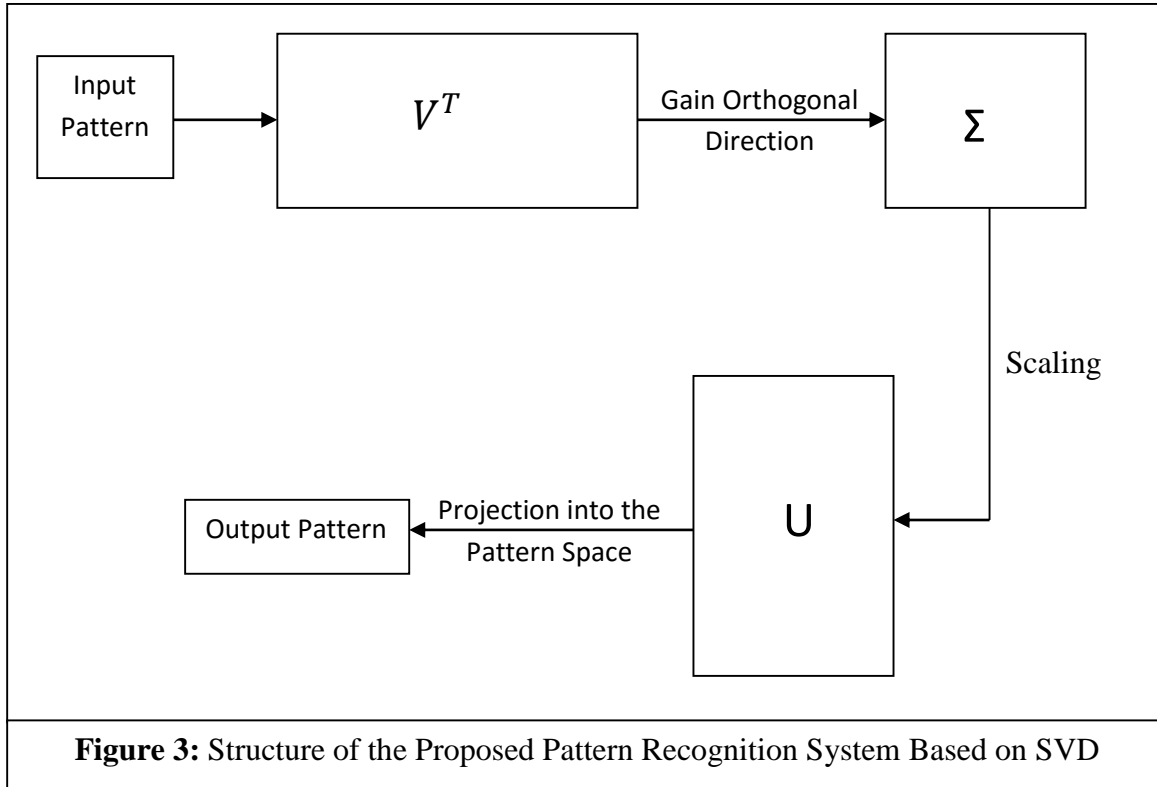


In this paper, SVD plays the main role in the pattern recognition where the proposed scheme is composed of two stages, as it is presented in figure (3):

$$A = U\Sigma V^T = \sum_{i=1}^N \sigma_i u_i v_i^T \quad \text{--- (3)}$$

We know from SVD that U represents the orthogonal basis for vector space of A and it is called left singular vectors; it represents the relationship among the attributes while V is the orthogonal basis for the relationship among patterns.

Semantic Pattern Recognition Based on Linear Algebra and Latent Semanti Analysis
Amjed Abbas Ahmed



The dataset has been selected with 100 different samples for the same character as it is shown in the following two examples:

Semantic Pattern Recognition Based on Linear Algebra and Latent Semanti Analysis
Amjed Abbas Ahmed

1	1	1	1	1
1	1	1	1	1

2	2	2	2	2
2	2	2	2	2

The orthogonal basis matrix can be interpreted as the original pattern and other patterns are the spanned vectors over the subspace constructed by the original pattern. What we are about in this paper is to build semantic pattern recognition system based on finding the basis vector for the given dataset.

Semantic space matrix is constructed according to the following formula:

$$entry_{i,j} = f(feature_i, pattern_j) \quad \text{--- (4)}$$

$$\forall pattern \ \forall feature \ IsAffectedBy(feature, pattern) \wedge \ HasValue(X \mid x \in R, feature) \quad \text{--- (5)}$$

$$\forall feature_i \ \forall feature_j \neq i \ IsCorrelated(feature_i, feature_j) \wedge \ HasCorrelationValue(X \mid x \in R, feature_i, feature_j) \quad \text{--- (6)}$$

All matrix entries should be affected by the correlation value; this is to add semantic to the vector composing each column within the semantic matrix.

**Semantic Pattern Recognition Based on Linear Algebra and
Latent Semanti Analysis
Amjed Abbas Ahmed**

In this paper the following features have been selected to represent patterns: {Mean, Std. Deviation, Median, Mode, Variance, Max, Min, and Correlation}. Table-1 represents the statistical calculations for hand script images for each pattern.

Table 1 : Statistical Calculations for Each Hand Script Images

feature	P_0	P_1	P_2	P_3	P_4	P_5	P_6
Mean	0.51020	0.57143	0.38776	0.53061	0.30612	0.57143	0.5231
Std. Deviation	0.50508	0.5	0.49229	0.50423	0.46566	0.5	0.4104
Median	1	1	0	1	0	1	1
Mode	1	1	0	0	0	1	0
Variance	0.2551	0.25	0.24235	0.25425	0.21684	0.25	0.2212
Max	1	1	1	1	1	1	1
Min	-1	-1	-1	-1	-1	-1	-1
Length	1.414	1.414	1.414	1.414	1.414	1.414	1.414

After subjecting the space constructed by table-1- to geometric analysis using the SVD and LSA we have the output shown in table-2 and table-3.

Semantic Pattern Recognition Based on Linear Algebra and Latent Semanti Analysis
Amjed Abbas Ahmed

Table 2 : Variance Matrix of the Attributes in Semantic Space

6.0756956	0	0	0	0	0	0
0	1.454168	0	0	0	0	0
0	0	0.8367016	0	0	0	0
0	0	0	0.08066494	0	0	0
0	0	0	0	0.06413336	0	0
0	0	0	0	0	0.008723614	0
0	0	0	0	0	0	3.6E-17

The maximum variance is given at the following direction

$$\begin{aligned}
 MaxVariance = & 0.4123798 X_1 - 0.360813 X_2 + 0.1807664 X_3 - 0.3317803 X_4 + \\
 & 0.4592246 X_5 - 0.5883154 X_6 + 1.49291E - 15X_7
 \end{aligned}$$

The analysis of the semantic space constructed by the above matrix is shown in the figure (4)

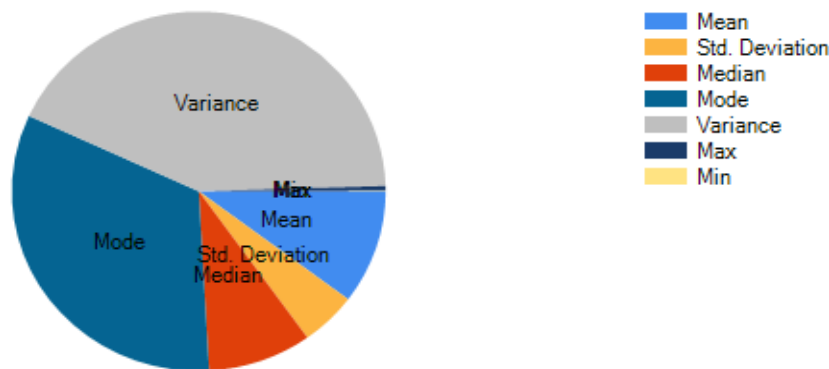


Figure 4: Analysis of Attribute Vector Space Using SVD and LSA

And according to the analysis in the space of attributes, the maximum variance can be captured by multiplying the first row of the matrix V by the variance matrix shown in table-2, the result is shown in figure (5).

**Semantic Pattern Recognition Based on Linear Algebra and
Latent Semanti Analysis
Amjed Abbas Ahmed**

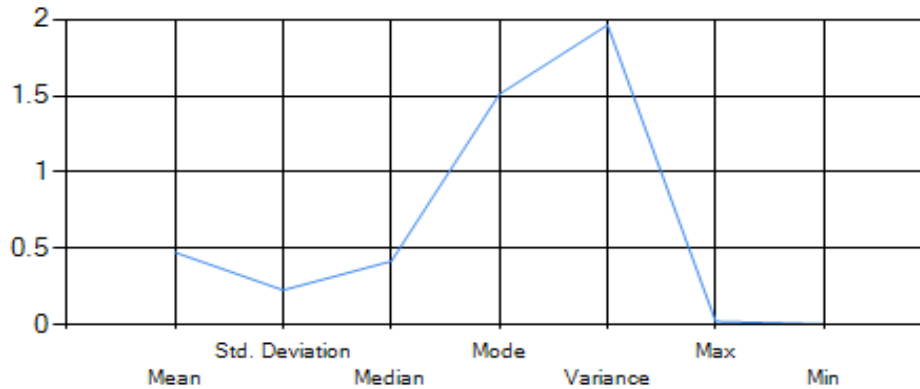


Figure 5: Weighting Each Attribute in Recognizing Hand scripts

Maximum recognition is captured by converging toward the following vector:

$$\text{Max Variance Vector} = 0.4722013\text{Mean} + 0.2245969 \text{ Std_Deviation} + 0.4140522 \text{ Median} + 1.509917 \text{ Mode} + 1.963358 \text{ Variance}.$$

Conclusions

From the results we concluded:

1. Applying SVD reduced the computation load due to cancel less affective components of the pattern and consider only the components that encapsulate the system dynamic.
2. Patterns are holding semantic relationships among their components through which these patterns can be recognized efficiently; LSA technique captures these semantic relationships and deploy it to recognize patterns with semantic domain.

Semantic Pattern Recognition Based on Linear Algebra and Latent Semanti Analysis Amjed Abbas Ahmed

References

- 1- Jucheng Yang, Yanbin Jiao, Jinfeng Yang, Zhijun Fang, Congcong Xiong, and Lei Shu, “Image Latent Semantic Analysis for Face Recognition”,
- 2- Matt Seigel, “Confidence Measures for Large Vocabulary Continuous Speech Recognition”, PhD Report, Department of Engineering, University of Cambridge, 2010.
- 3- O. Seredin, a. Kopylov, and V. Mottl, “Selection of Subsets of Ordered Features in Machine Learning”, 6th International Conference, MLDM 2009, Germany, 2009.
- 4- Cheng-Lin Liu, Hiroshi, and Hiromichi Fujisawa, “Performance Evaluation of Pattern Classifiers for Handwritten Character Recognition”, International Journal on Document Analysis and Recognition, Springer-Verla, Vol. 4, page 191-204, 2002.
- 5- Jianfeng Ren, Xudong Jiang, and Junsong Yuan, “Relaxed Local Ternary Pattern for Face Recognition”,
- 6- Koneru Anuradha and Manoj Kumar Tyagi, “A Novel Method of Face Recognition Using Lbp, Ltp and Gabor Features”, International Journal of Scientific & Technology Research, Vol. 1, Issue 5, 2012.
- 7- Thomas K. Landauer, Danielle S. McNamara, Simon Dennis and Walter Kintsch, “Handbook of Latent Semantic Analysis”, Lawrence Erlbaum Associates, Inc, USA, 2011, ISBN:978-1.4106-1534-3
- 8- NursuriatiJamil and SitiAisyahSa’adan, Proceeding, “Visual Informatics: Bridging Research and Practice”, visual informatics conference, IVIC 2009, Kuala Lumpur, Malaysia, 2009.