Spatial Linear Predictive Coding and its Error Matching for Signal Classification

TUAN D. PHAM[†,‡]
[†]Bioinformatics Applications Research Centre
[‡]School of Mathematics, Physics, and Information Technology
James Cook University
Townsville, QLD 4811
AUSTRALIA

Abstract: Mathematical analysis of the behavior of general dynamic systems based on linear prediction plays an essential role in many fields of science and engineering concerning the processing and representation of complex signals. This paper addresses the parameter estimation of the all-pole model of the linear predictive coding in the sense that the signal has both deterministic and random properties. Estimate of the model variance error is used as a basis for the derivation of a spatial distortion measure which can be used for matching spectral patterns.

Key–Words: Linear prediction, geostatistics, error matching, distortion measure, signal processing, classification.

1 Introduction

The theory of linear prediction still remains an important research area of digital signal processing, which has been originally applied for signal detection and signal coding in neurophysics, geophysics, and speech communication [1]. Research into time series analysis has been focused on the estimation of power spectra, cross spectra, coherence functions, autocorrelation and cross-correlation functions. Another active concern is the construction of a parametric model for modeling the behavior of some complex signal. The success of developing such model can guarantee its effective applications for solving many difficult problems in digital communications, engineering control, prediction and forcasting [1, 2, 3, 4].

In time series analysis, a continuous-time signal $s(t)$ is sampled to obtain a discrete-time signal $s(nT)$, where n is an integer variable and T is the sampling interval. For the sake of convenience, from now on we denote $s(nT)$ as $s(n)$ without the loss of generality.

The basic formulation of linear prediction is based on the assumption that a signal s_n is considered to be the output of some sytem with some unknown input u_n such that [1]

$$s_n = \sum_{k=1}^{p} a_k s(n-k) + G \sum_{l=0}^{q} b_l u(n-l) \qquad (1)$$

where $b_0 = 1$, the terms $\{a_k\}$ and the gain G are the parameters of the hypothesized system.

Equation (1) can be expressed in the frequency domain by taking the z transform on its both sides, which results in

$$H(z) = \frac{S(z)}{U(z)} = G \frac{1 + \sum_{l=1}^{q} b_l z^{-l}}{1 + \sum_{k=1}^{p} a_k z^{-k}} \qquad (2)$$

where $H(z)$ is the system transfer function, $U(z)$ is the z transform of $u(n)$, and $S(z)$ the z transform of $s(n)$ which is defined as

$$S(z) = \sum_{n=-\infty}^{\infty} s(n) z^{-n} \qquad (3)$$

The term $H(z)$ expressed in (2) is called the general *pole-zero* model – the roots of the numerator are the *zeros*, and the roots of the denominator are the *poles* of the model. Thus, two special cases of the model are the all-zero and all-pole models. As the names refer, for the all-zero model: $a_k = 0, \forall k$; whereas for the all-pole model: $b_l = 0, \forall l$. Our discussion is now focused on the all-pole model which assumes that the signal $s(n)$ can be determined as a linear combination of the past values and some input $u(n)$:

$$s(n) = \sum_{k=1}^{p} a_k s(n-k) + G u(n) \qquad (4)$$

where G is a gain factor and the transfer function $H(z)$ of the all-pole model simplyfies to

$$H(z) = \frac{G}{1 + \sum_{k=1}^{p} a_k z^{-k}} \qquad (5)$$

In general, the problem of the linear prediction based on the all-pole model is to determine the set of the predictor coefficients $\{a_k\}$ and the gain G. In many applications the input $u(n)$ is unknown. This fact further reduces the linear prediction model to

$$\hat{s}(n) = \sum_{k=1}^{p} a_k \, s(n-k) \tag{6}$$

where $\hat{s}(n)$ is the approximation of $s(n)$.

The prediction error $e(n)$ between the observed sample $s(n)$ and the predicted value $\hat{s}(n)$ can be defined as

$$e(n) = s(n) - \hat{s}(n) = s(n) - \sum_{k=1}^{p} a_k \, s(n-k) \tag{7}$$

Since the spectral properties of the signal can vary over time, the predictor coefficients at a given time n must be estimated from a short segment of the signal occuring around time n. Using the principle of least squares, we can find an optimal set of predictor coefficients by minimizing the mean-squared prediction error over a short segment of the whole signal.

A short-term signal, $s_n(m)$, and its error segment, $e_n(m)$, at time n can be defined as

$$s_n(m) = s(n+m) \tag{8}$$

and

$$e_n(m) = e(n+m) \tag{9}$$

The mean-squared error signal at time n to be minimized is defined as

$$E_n = \sum_{m} e_n^2(m) \tag{10}$$

which can be expressed in terms of $s_n(m)$ as follows.

$$E_n = \sum_{m} \left[s_n(m) - \sum_{k=1}^{p} a_k s_n(m-k) \right]^2 \tag{11}$$

Differentiating E_n, which is expressed in (11), with respect to each a_k and set the result to zero:

$$\frac{\partial E_n}{\partial a_k} = 0, \; k = 1, \ldots, p \tag{12}$$

giving

$$\sum_{m} s_n(m-i) s_n(m) = \sum_{k=1}^{p} a_k \sum_{m} s_n(m-i) s_n(m-k) \tag{13}$$

It can be noticed that the terms of the form $\sum s_n(m-i)s_n(m-k)$ are those of the short-term covariance of $s_n(m)$, that is

$$\phi_n(i,k) = \sum_{m} s_n(m-i) s_n(m-k) \tag{14}$$

One possible way of defining the limits on m expressed in (14) is to assume that the segment, $s_n(m)$, is zero outside the interval $0 \le m \le N-1$, where N is the size of the short segment. This assumption is equivalent to that the signal $s(m+n)$ is multiplied by a finite length window, $w(m)$, which zero outside the range $0 \le m \le N-1$. Thus the segment for minimization can be expressed as

$$s_n(m) = \begin{cases} s(m+n)\, w(m) & : \quad 0 \le m \le N-1 \\ 0 & : \quad \text{otherwise} \end{cases} \tag{15}$$

where $w(m)$ is usually a Hamming window.

Based on using the signal expressed in (15), the error signal $e_n(m)$ is exactly zero since $s_n(m) = 0$ for all $m < 0$, and for $m > N-1+p$ the prediction error is also zero because again $s_n(m) = 0$ for all $m > N-1$. Thus an optimal range of m used in defining the short segment of the sequence and the region over which the mean-squared error is minimized is from $m = 0$ to $m = N-1+p$ to minimize the errors at section boundaries. Using this range for m, the mean-squared error becomes [2]

$$E_n = \sum_{m=0}^{N-1+p} e_n^2(m) \tag{16}$$

and $\phi_n(i,k)$ can be rewritten as

$$\phi_n(i,k) = \sum_{m=0}^{N-1+p} s_n(m-i) s_n(m-k), 1 \le i \le p, 0 \le k \le p \tag{17}$$

or

$$\phi_n(i,k) = \sum_{m=0}^{N-1-(i-k)} s_n(m) s_n(m+i-k), 1 \le i \le p, 0 \le k \le p \tag{18}$$

Since (18) is a function of $(i-k)$, the covariance function $\phi_n(i,k)$ can be reduced to the simple autocorrelation function:

$$\phi_n(i,k) = r_n(i-k) = \sum_{m=0}^{N-1-(i-k)} s_n(m) s_n(m+i-k) \tag{19}$$

Since the autocorrelation function is symmetric, that is $r_n(-k) = r_n(k)$, the system of LPC equations can be expressed as

$$\sum_{k=1}^{p} r_n(|i-k|) a_k = r_n(i), \; 1 \le i \le p \tag{20}$$

which describes a set of p equations in p unknowns, and can be expressed in matrix form as

$$\mathbf{R}\,\mathbf{a} = \mathbf{r} \tag{21}$$

where $\mathbf{R}$ is a $p \times p$ autocorrelation matrix (Toeplitz matrix which is symmetric with all diagonal elements being equal), $\mathbf{r}$ is a $p \times 1$ autocorrelation vector, and $\mathbf{a}$ is a $p \times 1$ vector of prediction coefficients:

$$\mathbf{R} = \begin{bmatrix} r_n(0) & r_n(1) & r_n(2) & \cdots & r_n(p-1) \\ r_n(1) & r_n(0) & r_n(1) & \cdots & r_n(p-2) \\ r_n(2) & r_n(1) & r_n(0) & \cdots & r_n(p-3) \\ \cdot & \cdot & \cdot & \cdots & \cdot \\ r_n(p-1) & r_n(p-2) & r_n(p-3) & \cdots & r_n(0) \end{bmatrix}$$

$$\mathbf{a}^T = \begin{bmatrix} a_1 & a_2 & a_3 & \cdots & a_p \end{bmatrix}$$

and

$$\mathbf{r}^T = \begin{bmatrix} r_n(1) & r_n(2) & r_n(3) & \cdots & r_n(p) \end{bmatrix}$$

Thus, the LPC coefficients can be obtained by solving

$$\mathbf{a} = \mathbf{R}^{-1}\mathbf{r} \tag{22}$$

2 Spatial Linear Predictive Coding

Having outlined the theory of linear predictive coding (LPC), we present in this section a new approach for estimating the LPC model parameters based on the theory of regionalized variables [5] and the kriging estimation procedure [6, 7]. A regionalozed variable is thought to have characteristics intermediate between a random variable and a deterministic function - its values vary over space but are spatially correlated over some short distance. The degree of the spatial continuity of a regionalized variable can be expressed by a semivariogram (to be discussed later). We now describe how the theory of regionalized variables and the unbiased estimation of kriging can be used for modeling the all-pole linear prediction.

Consider a stationary random function that consists of several random variables, one for each of the available values and one for the unknown value. Let $V(s(n-k)), k = 1, \ldots, p$, be the random variables of $s(n-k), k = 1, \ldots p$, respectively. Let $V(s(n))$ be the random variable for $s(n)$. These random variables are assumed to have the same probability distribution, and the expected value of the random variables at all locations is $E\{V\}$. Thus, the estimate of $s(n)$ is also a random variable and expressed by a weighted linear combination of the random variables at p locations:

$$\hat{V}(s(n)) = \sum_{k=1}^{p} a_k \, V(s(n-k)) \tag{23}$$

And the error of estimation is

$$R(s(n)) = \hat{V}(s(n)) - V(s(n)) \tag{24}$$

Alternatively we have

$$R(s(n)) = \sum_{k=1}^{p} a_k V(s(n-k)) - V(s(n)) \tag{25}$$

The expected value of the error of estimate is

$$E\{R(s(n))\} = \sum_{k=1}^{p} a_k E\{V(s(n-k))\} - E\{V(s(n))\} \tag{26}$$

Based on the assumption that the random function is stationary, both $E\{V(s(n-k))\}$ and $E\{V(s(n))\}$ can be expressed as $E\{V\}$; thus (26) becomes

$$E\{R(s(n))\} = \sum_{k=1}^{p} a_k E\{V\} - E\{V\} \tag{27}$$

If the unbiased condition is imposed, then $E\{R(s(n))\}$ must be set to zero. Giving

$$E\{V\} \sum_{k=1}^{p} a_k = E\{V\} \tag{28}$$

resulting

$$\sum_{k=1}^{p} a_k = 1 \tag{29}$$

The variance of the random variable $V(s(n))$ which is the result of a weighted linear combination of other p random variables is given by

$$Var\{\sum_{k=1}^{p} a_k V(s(n-k))\} = \sum_{k=1}^{p}\sum_{j=1}^{p} a_k a_j Cov\{V(s(n-k))V(s(n-j))\} \tag{30}$$

Recalling that $R(s(n)) = \hat{V}(s(n)) - V(s(n))$ and using (30), the variance of the error can be expressed as either

$$\begin{aligned} Var\{R(s(n))\} &= Cov\{\hat{V}(s(n))\hat{V}(s(n)) - 2Cov\{\hat{V}(s(n))V(s(n))\} \\ &+ Cov\{V(s(n))V(s(n))\} \end{aligned} \tag{31}$$

or

$$\sigma_R^2 = \sigma^2 + \sum_{k=1}^{p}\sum_{j=1}^{p} a_k a_j C_{kj} - 2\sum_{k=1}^{p} a_k C_k \tag{32}$$

which defines the variance of error as a function of $a_1, \ldots, a_p$.

An optimal choice for the predictor parameters $a_1, \ldots, a_p$ is to minimize σ_R^2. Introducing a Lagrange multiplier β into (32) we have

$$\sigma_R^2 = \sigma^2 + \sum_{k=1}^{p}\sum_{j=1}^{p} a_k a_j C_{kj} - 2\sum_{k=1}^{p} a_k C_k + 2\beta(\sum_{k=1}^{p} a_k - 1) \tag{33}$$

The error variance term, σ_R^2, can now be minimized by differentiating (33) with respect to the predictor coefficients and the Lagrange parameter, and setting each one to zero. By doing so, we obtain the following equations.

$$\sum_{j=1}^{p} a_j C_{kj} + \beta = C_{kn},\ \forall k = 1, \ldots, p. \tag{34}$$

$$\sum_{k=1}^{p} a_k = 1 \tag{35}$$

The above system of equations are known as the ordinary kriging system [], which can be expressed in matrix notation as

$$\mathbf{C}\,\mathbf{a} = \mathbf{D} \tag{36}$$

where

$$\mathbf{C} = \begin{bmatrix} C_{11} & \cdots & C_{1p} & 1 \\ \cdot & \cdots & \cdot & \cdot \\ \cdot & \cdots & \cdot & \cdot \\ \cdot & \cdots & \cdot & \cdot \\ C_{p1} & \cdots & C_{pp} & 1 \\ 1 & \cdots & 1 & 0 \end{bmatrix}$$

$$\mathbf{a} = \begin{bmatrix} a_k & \cdots & a_p & \beta \end{bmatrix}^T$$

$$\mathbf{D} = \begin{bmatrix} C_{1n} & \cdots & C_{pn} & 1 \end{bmatrix}^T$$

Thus the values of the spatial predictor coefficients can be obtained by solving

$$\mathbf{a} = \mathbf{C}^{-1}\,\mathbf{D} \tag{37}$$

The sample covariance used for the kriging estimator can be calculated as

$$C(h) = \frac{1}{N(h)} \sum_{(i,j)|h_{ij}=h} s(j) - (\frac{1}{n}\sum_{k=1}^{n} s(k))^2 \tag{38}$$

in which the sample covariance is a function of the lag distance h, $N(h)$ is the number of pairs that $s(i)$ and $s(j)$ are separated by h, and n is the total number of data.

On the derivation of the error of variance, it is assumed that the random variables have the same mean and variance which lead to the development of the mathematical relationship between the variogram, denoted as $\gamma(h)$, and the covariance [7]

$$\gamma(h) = \sigma^2 - C(h) \tag{39}$$

where the sample $\gamma(h)$ is defined as

$$\gamma(h) = \frac{1}{2N(h)} \sum_{(i,j)|h_{ij}=h} [s(i) - s(j)]^2 \tag{40}$$

Using the variogram, the kriging weights can be determined by solving

$$\sum_{k=1}^{p} a_k \gamma_{jk} - \beta = \gamma_{jn},\ j = 1, \ldots, p \tag{41}$$

and

$$\sum_{k=1}^{p} a_k = 1 \tag{42}$$

The variance of the estimation residual (error) can readily be determined by

$$\sigma_R^2 = \mathbf{a}^T \mathbf{D} \tag{43}$$

Taking the square root of (43) gives the standard error of the estimate:

$$\sigma_R = (\mathbf{a}^T \mathbf{D})^{\frac{1}{2}} \tag{44}$$

3 LPC-VQ-based Decision Logic for Signal Classification

To apply the results obtained from the linear prediction for classification of unknown signals, the method of vector quantization can be utilized to generate a decision logic for classification. We discuss herein the implementation of both conventional and spatial distortion measures for the VQ-codebook design.

A distortion measure between two vectors $\mathbf{x}$ and $\mathbf{y}$, denoted as $D(\mathbf{x}, \mathbf{y})$, is considered to be a cost of reproducing any input vector $\mathbf{x}$ as a reproduction of vector $\mathbf{y}$. Given such a distortion measure, the mismatch between two signals can be quantified by an average distortion between the input and the final reproduction. Intuitively, a match of the two patterns is good if the average distortion is small.

A popular distortion measure is the likelihood ratio (LR) distortion. The LR distortion measure, D_{LR}, is defined by [2]

$$D_{LR} = \frac{\mathbf{a'}^T\,\mathbf{R}_s\,\mathbf{a'}}{\mathbf{a}^T\,\mathbf{R}_s\,\mathbf{a}} - 1 \tag{45}$$

where $\mathbf{R}_s$ is the autocorrelation matrix of signal s associated with its LPC coefficient vector $\mathbf{a}$, and $\mathbf{a}'$ is the LPC coefficient vector of signal s'.

Based on the same principle derived for the likelihood ratio distortion and using (43), the spatial distortion, denoted as D_S, can be defined as

$$D_S = \frac{\mathbf{a}^T\mathbf{D}}{\mathbf{a'}^T\mathbf{D}} - 1 \tag{46}$$

where $\mathbf{a}$ defined in (36) is the spatial LPC vector of signal s, $\mathbf{D}$ is the matrix defined in (36) associated with s , and $\mathbf{a}'$ is the spatial LPC vector of signal s'.

Now assume we have a set of T frames or subsequences of the whole sequence, which are represented by the corresponding set of T LPC vectors $\mathbf{A} = \{\mathbf{a}_1, \mathbf{a}_2, \ldots, \mathbf{a}_T\}$, where $\mathbf{a}_t = (a_{t1}, a_{t2}, \ldots, a_{tp})$. It can be seen that these LPC vectors represent a type of feature of the sequence. Let the codebook of the LPC vectors be $\mathbf{C} = \{\mathbf{c}_1, \mathbf{c}_2, \ldots, \mathbf{c}_N\}$, where $\mathbf{c}_n = (c_{n1}, c_{n2}, \ldots, c_{np})$, $n = 1, 2, \ldots, N$ are codewords. Each codeword $\mathbf{c}_n$ is assigned to an encoding region R_n in the partition $\Omega = \{R_1, R_2, \ldots, R_N\}$. The source LPC vector $\mathbf{a}_t$ can be represented by the encoding region R_n and expressed by

$$V(\mathbf{a}_t) = \mathbf{c}_n, \text{ if } \mathbf{a}_t \in R_n \tag{47}$$

The main idea of LPC based vector quantization (VQ) is to find an optimal codebook such that for a given training set $\mathbf{A}$ and a codebook size N, the average distortion in representing each LPC vector $\mathbf{a}_t$ by the closest codeword $\mathbf{c}_n$ is minimum. In mathematical terms we express

$$D^* = \min_{\mathbf{c}_n} \left[\frac{1}{T} \sum_{t=1}^{T} \min_{1 \le n \le N} (D(\mathbf{c}_n, \mathbf{a}_t)) \right] \tag{48}$$

where D is an LPC distortion and D^* is the average distortion of the vector quantizer.

In general, a vector quantizer is a system that maps a sequence of continuous or discrete vectors into a digital sequence suitable for storage in a digital channel. Vector quantization has been found to be very useful for encoding LPC vectors [8]. In other words, LPC vectors coupling with vector quantization have found to be very effective for signal coding and recognition [9]. Although there are several data partioning methods for the determination of an optimal VQ codebook. One of the most popular methods for VQ is the LBG (Linde, Buzo and Gray) algorithm [10]. The LGB-VQ method requires an initial codebook, and iteratively bi-partitions the codevectors based on the optimality criteria of nearest-neighbor and centroid conditions until the number of codevectors is reached.

The classification system based on the LPC analysis and VQ codebook approach works as follows. The given input signal is analyzed by the LPC giving the sequence of LPC vectors. The resultant LPC vectors are quantized using the number of codebooks according to the number of different classes. The distortions with respect to each codebook are accumulated across the whole test. The average spectral distortion (dissimilarity) measure between an unknown sample and a particular known class is

$$\bar{D}(\mathbf{x}_m, \mathbf{c}^i) = \frac{1}{T} \sum_{m=1}^{T} \min_{1 \le j \le J} D(\mathbf{x}_m, \mathbf{c}_j^i) \tag{49}$$

where D is a spectral distortion measure, $\bar{D}$ is the average distortion, $\mathbf{x}_m$ is an LPC vector of the unknown signal, T is the number of LPC vectors of the unknown signal, $\mathbf{c}_j^i$ is the j LPC-VQ codevector of a particular class represented by codebook $\mathbf{C}^i$, and J the size of $\mathbf{C}^i$.

The unknown signal is assigned to class i^* if the average distortion measure of its LPC feature vector $\mathbf{x}_m$ and the LPC feature codebook $\mathbf{C}^i$ is minimum, that is

$$i^* = \arg\min_i \bar{D}(\mathbf{x}_m, \mathbf{C}^i) \tag{50}$$

Table 1: k-fold cross validation results for ovarian cancer data (μ_{cl}: control mean, μ_{cr}: cancer mean)

	SVM		LPC-VQ/Spatial LPC-VQ	
k	μ_{cl}	μ_{cr}	μ_{cl}	μ_{cr}
2	0.8930	0.9492	0.9224/0.9231	0.9637/0.9640
4	0.9058	0.9722	0.9327/0.9320	0.9811/0.9821
6	0.9094	0.9760	0.9348/0.9332	0.9825/0.9814
8	0.9098	0.9784	0.9362/0.9359	0.9852/0.9887
10	0.9096	0.9801	0.9377/0.9412	0.9885/0.9890

4 Application

The identification of biomarkers using mass spectrometry (MS) data [11, 12, 13, 14] is a challenging task which requires the combination of the contrast fields of knowledge of modern biology, signal processing, and pattern recognition. The basic problem is to classify an unknown MS signal as either the control (non-disease) group or the disease group.

The proposed method (spatial LPC-VQ) was tested using a public ovarian high-resolution SELDI-TOF mass spectrometry dataset. Regarding the implementation of proposed method, the number of poles $p = 8$ was specified for the LPC analysis. The codebook size of 64 codevectors was used to generate the prototype for the control and cancer classes. Each MS sequence was split into multiple frames of 150 data points having 20 points overlapping between the two adjacent frames. At present, these paramters for signal processing were arbitrarily chosen and based on the experience that these values have been considered reasonable for the classification of speech signals. These paramters were also used for the conventional (non-spatial) LPC-VQ method.

The ovarian high-resolution SELDI-TOF mass spectrometry dataset, which can be obtained from the FDA-NCI Clinical Proteomics Program Databank (http://home.ccr.cancer.gov/ncifdaproteomics/ppatterns.asp), was used to test the proposed spatial LPC-VQ based method. The dataset was generated using a non-randomized study set of ovarian cancers and control specimens on an ABI Qstar fitted with a SELDI-TOF source to study ovarian cancer case versus high-risk control. The dataset consists of 100 control samples and 170 cancer samples. The validation of the classification of the proposed approach was designed with similar strategies

to those carried out in [15], who applied support vector machine (SVM) for the classification, so that comparisons can be made. The measure of performance is the k-fold cross validation where k = 2, 4, 8, and 10, and each k-fold validation was carried out 1000 times.

It is noted that the raw ovarian high-resolution SELDI-TOF dataset used by Yu *et al.* [15] consists of 95 control samples and 121 cancer samples; while the raw ovarian high-resolution SELDI-TOF dataset we used to test the performance of the proposed approach has 100 control samples and 170 cancer samples.

Table 1 shows the mean classification values of the k-fold cross validation obtained from the SVM classification on the preprocessed data and the proposed method. The overall results obtained for the three types of validation of the classification performance show that both the spatial and non-spatial LPC-VQ methods are more favorable for MS-based ovarian cancer identification than the other approach in two folds - feature extraction and classification methods. The spatial and conventional LPC-VQ are competitive and can be complementary for information fusion which may help improve the classification rate.

In terms of feature extraction, the procedure for extracting the LPC-based features is more straightforward than the other feature extraction method [15] which transforms raw MS data to binned MS data, from binned MS data to Kolmogorov-Smirnov(KS)-test based feature selection, from KS-test based features to the restriction of coefficient of variation (CV), and finally from the restriction of CV to wavelet coefficients. In terms of pattern classification, the classification using the VQ-based decision rule is much simpler than the SVM-based classifier; however it is not meant that the SVM-based classifier is inferior to the VQ-based decision rule for classifying MS data. The LPC-based coefficients can be used as robust features for training different classifiers to improve the performance capable of more effective discrimination of complex patterns.

5 Conclusion

We have presented and discussed a new approach for estimating the spatial linear predictor coefficients using the principles of the theory of regionalized variables and the unbiased kriging estimator. Based on the spatial LPC vectors and the error variance of kriging estimation, a novel spatial distortion measure has been derived as a new formulation for error matching between two signals. Both spatial LPC-VQ and non-spatial LPC-VQ methods have been applied to classify control and ovarian cancer classes using mass spectrometry data. The experimental results have shown the promising applications of both methods whose results can be combined to improve the classification rates [9], and potential applications for other related problems in bioinformatics [16].

References:

[1] J. Makhoul, Linear prediction: a tutorial review, *Proc. IEEE*, 63 (1975) 561-580.

[2] L. Rabiner, and B.H. Juang, *Fundamentals of Speech Recognition*. New Jersey, Prentice Hall, 1993.

[3] Ingle, V.K., and Proakis, J.G., *Digital Signal Processing Using Matlab V.4*. Boston, PWS Publishing, 1997.

[4] T.D. Pham and M. Wagner, A geostatistical model for linear prediction analysis of speech, *Pattern Recognition*, 31 (1998) 1981-1991.

[5] G. Matheron, La théorie des variables régionalisées at ses appplications, *Cahier du Centre de Morphologie Mathématique de Fontainebleau. É*cole des Mines, Paris, 1970.

[6] A.G. Journel, and C.J. Huibregts, *Mining Geostatistics*. Academic Press, London, 1978.

[7] E.H. Isaaks, and R.M. Srivastava, *An Introduction to Applied Geostatistics*. Oxford University Press, New York, 1989.

[8] R.M. Gray, Vector quantization, *IEEE ASSP Mag.*, 1 (1984) 4-29.

[9] Rabiner, L.R., Sondhi M.M., and Levinson, S.E., A vector quantizer incorporating both LPC shape and energy, *Proc. 1984 Int. Conf. Acoustics, Speech, and Signal Processing*, pp. 17.1.1-17.1.4, 1984.

[10] Linde, Y., Buzo, A., and Gray, R.M., An Algorithm for Vector Quantization", *IEEE Trans. Communications*, **28** (1980) 84-95.

[11] R. Aebersold, and M. Mann, Mass spectrometry-based proteomics, *Nature*, 422 (2003) 198-207.

[12] E.F. Petricoin *et al.*, Use of proteomic patterns in serum to identify ovarian cancer, *Lancet*, 359 (2002) 572-577.

[13] T.P. Conrads, M. Zhou, E.F. Petricoin III, L. Liotta, and T.D. Veenstra, Cancer diagnosis using proteomic patterns, *Expert Rev. Mol. Diagn.*, 3 (2003) 411-420.

[14] J.D. Wulfkuhle, L.A. Liotta, and E.F. Petricoin, Proteomic applications for the early detection of cancer, *Nature*, 3 (2003) 267-275.

[15] J.S. Yu, S. Ongarello, R. Fiedler, X.W. Chen, G. Toffolo, C. Cobelli, and Z. Trajanoski, Ovarian cancer identification based on dimensionality reduction for high-throughput mass spectrometry data, *Bioinformatics*, 21 (2005) 2200-2209.

[16] T.D. Pham, LPC cepstral distortion measure for protein sequence comparison, *IEEE Trans. NanoBioscience*, 5 (2006) 83-88.