International Journal of Engineering & Technology, 7 (2.17) (2018) 116-123

International Journal of Engineering & Technology

Website: www.sciencepubco.com/index.php/IJET

Research Paper

Efficient Adaptive Exon Prediction for DNA study using Proportionate LMS Variants

Srinivasareddy Putluri, Md Zia Ur Rahman*

Department of Electronics and Communication Engineering, Koneru Lakshmaiah Education Foundation, Green fields, Guntur DT, AP, 522502, India.
**Corresponding author E-mail:mdzr55@gmail.com*

Abstract

In the field of Bio-informatics, locating the exon fragments in a deoxyribonucleic acid (DNA) sequence is an important and vital work. Study of protein coding regions is a wide phenomenon in identification of diseases and design of drugs. The regions of DNA that have the protein coding information are termed as exons. Hence identifying the exon segments in a genomic sequence is a crucial job in bio-informatics. Three base periodicity (TBP) has been observed in the regions of DNA sequences can be easily determined by applying signal processing methods. Adaptive signal processing techniques found to be useful than other available methods. This is due to their unique capability to alter weight coefficients based on genomic sequence. We propose efficient adaptive exon predictors (AEPs) based on these considerations using Proportionate Normalized LMS (PNLMS) algorithm and Maximum Proportionate Normalized LMS (MPNLMS) algorithm to improve exon locating ability and better convergence. To ease the complexity of computations in the denominator during filtering process, proposed AEPs using PNLMS and its maximum variants are combined with signature algorithms. Hybrid variants of proposed AEPs include PNLMS, DCPNLMS, ECPNLMS, SSPNLMS, MPNLMS, MDCPNLMS, MECPNLMS and MSSPNLMS algorithms. It was shown that the AEP based on MDCPNLMS is superior in applications of exon identification depending on performance measures with Sensitivity 0.7346, Specificity 0.7483 and precision 0.7325 for a genomic sequence with accession AF009962 at a threshold of 0.8. Finally the capability of several AEPs in predicting exon locations is verified using different DNA sequences found in National Center for Biotechnology Information (NCBI) gene database.

Keywords: *adaptive exon predictor; bioinformatics; computational complexity; deoxyribonucleic acid; three base periodicity; sensitivity; specificity; precision.*

1. Introduction

A substantial purpose of research in the field of bio-informatics is to study the nature of information along with its role in learning about a particular job encoded by the gene. An essential step to attain this goal is tracing the protein coding areas in a gene sequence [1]. Identification of exon sections is an mammoth space of exploration in bio-informatics. A division is frmed by crucial gene sequences in organisms that are required for the fertility, growth, or persistence [2]. As a result, identifying the protein coding sections is not only interesting, but also has realistic value to find out human diseases [3] and determine drug targets in new pathogens [4] - [6]. The intragenic and genic sections are available in a DNA sequence. In bio-informatics, the subarea which emphases on finding the exon sections in a DNA sequence is acknowledged as exon identification. Learning of principal exon segment structure aids in study of the ancillary and tertiary assembly of exon segments. All anomalies can be identified, can design drugs and treat diseases once the complete structure of protein coding sections is evaluated. The learnings help to know the phylogenic trees evaluation[7] - [8]. At present, a fast development of raw data of genomic sequences needs useful biological elucidations, but more cost is implicated to perform biological experiments for predicting gene locations and there is still a practical demand for efficient and fast tools mainly to identify genes, to study sequences and know their functions [9] - [10]. Whole bodily entities are alienated into two segments, named as eukaryotes plus prokaryotes based on the fundamental molecular cell structure. The protein coding segments accountable for coding of protein sectors are uninterrupted and elongated in archaea; prokaryotes and bacteria are the examples of prokaryotes. The genes are the consolidation of exon segments alienated by means of lengthy segments which are not involved in coding in eukaryotes [11]. The segments that are liable for protein coding are named as exons, whereas the sections not involved in coding are called as introns. Whole living organisms other than archaea and bacteria, all originate beneath this category. The exon segments found in eukaryotes of human beings are merely 3% of the gene sequence and the residual 97% are sections that are not involved in coding of protein. Henceforth finding the regions those involved in coding of protein is a precarious job [12] - [13]. The protein coding sections in almost whole DNA sequences will exhibit the three base periodicity (TBP). In the plot for power spectral density (PSD), a strident peak is observed at a frequency ‘f’ equal to 1/3 [14]. Various prediction techniques for exons are prevailing in works built on several techniques of signal processing [15] - [19]. Nevertheless, the size of the actual genomic sequence is remarkably long and changes between sequences. To process such genomic sequences, adaptive techniques are found to be vital methods. The property of three base periodicity is beneficial to determine the protein coding segments in a DNA sequence [20]. Adaptive methods are preferred for very long sequence processing in numerous repititions that can modify coefficients of weight in agreement to the numerical actions of input gene sequence. Here, we propose to develop an Adaptive Exon Predictor (AEP) using adaptive methods by using proportionate normalization concept in which the matrix for gain

Copyright © 2018 Authors. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

essentially assessment of taps in contrast to the amount during normalization. The Least mean squares (LMS) technique is an acceptable vital adaptive method. This technique is pervasive because it is easy for execution. Glitches such as amplification of gradient noise, drift in the weight and deprived convergence are agonized by this algorithm. Also, the speed of convergence is sluggish while the Eigen value extent is added; moreover lesser is the performance for low SNR. To daze the stability problems, innumerable Normalized LMS (NLMS) techniques are recommended. Superior performance of MSE with controlled step size and independent of signal power is the advantage of the normalized. Therefore, to further upsurge the performance of AEP, we state to use proportionate normalized LMS adaptive algorithm with its signed variants. The four resultant algorithms are Proportionate Normalized LMS (PNLMS), Data Clipped Proportionate Normalized LMS (DCPNLMS), Error Clipped Proportionate Normalized LMS (ECPNLMS) and Sign Sign Proportionate Normalized LMS (SSPNLMS) algorithms. Here, Proportionate Normalized LMS (PNLMS) is used for filtering. Extracting the sparse coefficients and weighing suitably is the evident benefit. Among the numerous NLMS algorithms, this acts as one of the best technique as this lessens the spread of eigenvalue that clues to quicker convergence. PNLMS is analogous to the proposed NLMS in [16], with respect to normalization. With regard to the gain matrix, variance is evident and principally taps are weighed based on their magnitude. In regards to the analysis offered in [17], this is vibrant that the PNLMS enjoys stability as NLMS and it also upsurges the rate of convergence thru weighing of the lethargic taps by lower weight. Normalized version of Proportionate LMS is called as Proportionate Normalized LMS (PNLMS) algorithm. PNLMS algorithm and its signed variants overcome the difficulties of LMS and mend ability of detecting the exon segments and provide superior convergence performance. It lessens mean square error (MSE) in the progression of exon identification. The complexity in computations for an adaptive technique shows a vivacious role.

Essentially, while the genomic sequence length is very huge, if the signal processing method has added computational complexity, the samples join at the AEP input. This causes inter symbol interference (ISI) and leads to inexactness in the identification of exons. Moreover, when the projected AEPs are implemented on nano device or VLSI circuit, the added computational complexity inclines to bigger size of circuit and additional actions. Henceforth, to handle the complexity in computations of proposed method in actual solicitations the techniques of adaptive nature are associated with techniques of signum function. Signum function is applied in sign algoritms and the number of operations for multiplication is reduced in the denominator [21]. The three streamlined signum methods are signum based regressor algorithm (SRA), signum algorithm (SA) and signum signum algorithm (SSA). Therefore, in order to lessen the computational complexity we conglomerate the three signum algorithms with the normalized LMS algorithm. Due to normalization in these algorithms, the denominator of the weight update equation has to calculate multiplications equal to the numeric value of tap length of the algorithm. When the tap length is higher, which is mutual in real time applications the large tap length origins an additional computational affliction on the AEP. This is minimized to one, regardless of extent os the tap expending a method so-called maximum normalization [22].The consequential maximum normalized versions of proposed AEPs are Maximum Proportionate Normalized LMS (MPNLMS), Maximum Data Clipped Proportionate Normalized LMS (MDCPNLMS), Maximum Error Clipped Proportionate Normalized LMS (MECPNLMS), and Maximum Sign Sign Proportionate Normalized LMS (MSSPNLMS) algorithms[16] - [19]. In the normalization version of the PLMS algorithm the connection amongst the reference input and error is normalized thru a factor equivalent to square of the norm. In normalized and maximum normalized algorithms, the gradient noise application problem is minimized and they converge more rapidly than the conventional LMS algorithm. Henceforward, PNLMS technique has a steady state error and convergence rate improved than LMS method [23]. Based on proposed proportionate normalized LMS and its maximum proportionate normalized variants, many AEPs are developed and verified the performance with actual gene sequences attained commencing thru the databank of National Center for Biotechnology Information (NCBI) [24]. We consider convergence characteristics, sensitivity (Sn), specificity (Sp) and precision (Pr) as metrics to assess the performance of the numerous recommended AEP techniques. Techniques with adaptive nature, outcomes of proposed methods and debate proceeding the performance of a number of proposed techniques is illustracted in subsequent units of the proposed work. Several adaptive signal processing techniues are presented in [25]-[50].

2. Adaptive Algorithms for Exon Prediction

Here, the DNA sequence input is altered into digital representation. It acts as a crucial chore in processing of gene sequences since such procedures are realistic solitary with discrete or digital signals. At present, the numerical mapping is inured of transforming input gene signal into digital data [14]. The method used for mapping is accustomed to exemplify an gene input sequence as four numerical sequences. By means of binary mapping, the happening of nucleotide by a location is specified by 1 and nonexistence as 0. Currently the ensuing sequence is proper for an adaptive algorithm as a input. Four numerical gene sequences are used as adaptive filter input [15]. At present, proposed AEP based technique is functional for renovating numerical signals. For instance I(n) be the input gene signal, mapping digital signal be M(n), T(n) be the confirmed gene signal based on TBP behavior, adaptive algorithm output be Y(n) and the signal feedback for apprising weight coefficients of the algorithm be F(n). Deliberate an adaptive LMS technique with a length equal to 'R'. The succeeding coefficient of weight can be forecasted depending on the present coefficient of weight, parameter for step size 'S', sequence input sample I(n) at the moment and the signal for feedback D(n) spawned in the loop of feedback in the presented technique. Algebraic relation along with the study of LMS technqiue was explained in [23]. The block characteristic illustration for probable AEP based technique is presented in Figure 1.

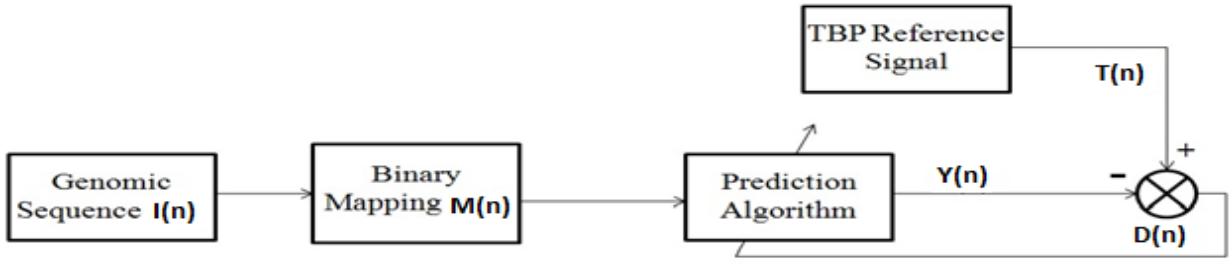


Fig. 1: Block diagram for adaptive exon predictor.

Owing to the ease and sturdiness, conservative LMS technique might be used in applications of exon identification. In lieu of stability along with convergence, filter of LMS requires former learning of level of power input to choose the parameter of step size. As it is typically one of the unknowns statistically, it is usually probable afore start of process for adaptation depending on data. On the other hand LMS method vacillates two downsides in real situations. This is real that the vector for data input is proportional to apprise way for weight unswervingly, thru witnessing the recursion of weight updation of LMS technique. Other problem is fixed step size. In real, a technqiue is deliberate in a way that, it will knob both resilient and feeble signals. Henceforth, tap coefficients must adjust accordingly based on the input and output variations of filter. As a result, LMS technique undergoes a amplification problem of gradient noise, while the data input vector is big. Normalization has to be applied to dodge this problem. Herewith, the attuned filter vector coefficient for weight with regard to Euclidian vector input with squared norm is normalized at each iteration.

The update equation for weight of adaptive LMS technique is given by

$$u(n+1) = u(n) + S\, I(n)D(n) \tag{1}$$

Low complexity in computations for technique with adaptive nature is extremely required as part of applications related to exon identification to develop devices at a nano scale. Decline is mostly gettable thru snipping whichever the data for input or response sequence or together. Techniques built on error snipping for data is illustrated in [20]. These are signum regressor technique (SRA), signum technique (SA) and signum signum technique (SSA). Among the adaptive algorithms, the signum methods has a rate of merging along with an error with stable state which is marginally lower compared to LMS method for the same setting of parameters. The function using signum is written as follows.

$$C\{D(n)\} = \begin{Bmatrix} 1: D(n) > 0 \\ 0: D(n) = 0 \\ -1: D(n) < 0 \end{Bmatrix} \tag{2}$$

To reduce the complexity in computations compared with LMS adaptive algorithms, we use SRA, SA and SA adaptive algorithms. The computational complexity of signum algorithms is more lower than the LMS technqiue. The Data Clipped LMS (DCLMS) technique is attained from the conventional recursion of LMS by changing the tap vector input I(n) by the vector C[D(n)], where signum function C is pragmatic to the vector D(n) on element basis. This is also called as clipped LMS as we are clipping the input signal data.

The update equation for weight of adaptive CLMS technique is reperesented by

$$u(n+1) = u(n) + S\, I(n)C[D(n)] \tag{3}$$

The update equation for weight of adaptive ECLMS technique is obtained by replacing I(n) with its signed form and is given by

$$u(n+1) = u(n) + S\, C[I(n)]D(n) \tag{4}$$

Similarly, the weight update relation of DECLMS algorithm is obtained by replacing I(n), D(n) with its signed forms and is given by

$$u(n+1) = u(n) + S\, C[I(n)]C[D(n)] \tag{5}$$

To daze the amplification problem of gradient noise associated with the filter of conventional LMS, the normalized filter of LMS presents a own problem, specifically the small input tap vector I(n), hitches may arise numerically which can be mitigated by dividing a lesser value using a norm in squared form. The above recursion can be modified by totaling a low positive constant ε, to overcome this difficulty. The parameter ε is chosen in a way to avoid much smaller value in denominator and much bigger value for step size parameter.

Now the step size parameter is written as,

$$S(n) = \frac{S}{\varepsilon + ||I(n)||2} \tag{6}$$

where S(n) is a step size of normalized nature using 0 <S< 2. Changing S in the update expression of LMS for weight vector of S(n) results to the DNLMS, which is given as

$$u(n+1) = u(n) + \frac{S}{||I(n)||2} I(n).D(n) \tag{7}$$

The denominator of the equation is made to control the convergence with the squared regressor term. This provides the algorithm stability against the signal power. The term δ is unaccustomed for sidestep the problems of stability when sequence contains the zero valued coefficients. Also, this behaves as the variant form of LMS due to the scaling of the step size and this improves the convergence. The constant in the denominator is introduced to prevent the algorithm to become unstable when the squared term tends to become zero. The above algorithm gives the reduced error, but the squared term in the denominator will increase the number of MAC operations, this increase the complexity and time to converge will increase. To reduce the number of computations in [16] and to feat the sparsity prevailing in data thru appropriately choosing the taps and weighing on distinct basis is proposed to update only required tap coefficients rather than all taps of the filter. Weighing is done through use of gain matrix, G. This yields a proportionate normalized LMS (PNLMS) algorithm.

Now the tap update equation of PNLMS is mathematically represented as,

$$u(n+1) = u(n) + \frac{SG}{\varepsilon + (I(n))^T G\, I(n)} I(n)D(n) \tag{8}$$

where G is the Gain matrix. Thus the alteration amid PNLMS and NLMS is witnessed in rapports of matrix for gain in the fraction part from the above equation. Pertaining to power of signal and a trivial co-efficient called factor of leakage based on normalization, ε that is significant to circumvent the stability glitch conditioned with signal power attains insignificant value. The above procedure reduces the computations involved but not the complexity. To reduce the complexity in this paper the sign algorithms are introduced. These algorithms have less convergence compared to DNLMS but the complexity reduces and the error will be little high. The signum function is successfully used in [17]-[19] for accurate prediction of exon location and better convergence. The sign algorithms are of three types, namely sign regressor, sign and sign sign algorithms. Therefore, in order to curtail the complexity in computations of PNLMS algorithm, PNLMS is combined with sign based techniques. Hybrid versions are named as MNSRLMS, MNSLMS and MNSSLMS algorithms.

The update equations for weight of adaptive DCPNLMS, ECPNLMS and SSPNLMS algorithms are written as,

$$u(n+1) = u(n) + \frac{SG}{\varepsilon + (I(n))^T G\, I(n)} C[I(n)]\, [D(n)] \tag{9}$$

$$u(n+1) = u(n) + \frac{SG}{\varepsilon + (I(n))^T G\, I(n)} [I(n)]\, C[D(n)] \quad (10)$$

$$u(n+1) = u(n) + \frac{SG}{\varepsilon + (I(n))^T G\, I(n)} C[I(n)]\, C[D(n)] \quad (11)$$

In the equations (8)-(11) the denominator of the normalization function requires "R" multiplications. When the filter length is large, the normalization function requires many multiplications. To avoid these excess multiplications we propose a normalization phenomenon, in which only the maximum value of input block is utilized for normalization. Using this maximum normalized approach only one multiplication is needed instead of "R" multiplications. Using the algorithms mentioned above, computational complexity is equivalent to "R" MACs is lowered towards single multiply and accumulate operation in the denominator by regulating size of step by means of an input data vector having maximum value. The version is called as a maximum proportionate normalized LMS (MPNLMS) algorithm.
The weight update equations of these maximum proportionate normalized algorithm is given as,

3. Computational Complexity and Convergence Issues

In general, to compare and estimate algorithm complexity, number of multiplications required to complete the operation is taken as a measure. However, most of the DSP's have a built in hardware support for multiplication and accumulation (MAC) operations. Usually they perform this operation in a single instruction cycle as well as addition or subtraction. In this paper, we are not trying to provide an exact analysis of a computational complexity; rather we concentrate on presenting a comparison between different adaptive algorithms. The computational complexity figures required in computation of various algorithms considered are summarized in the Table 1. Furthermore, these algorithms provide an elegant means for adaptive exon prediction applications, for instance the projected signum algorithms are chiefly at liberty from operation of multiplication. For example, LMS technique entails R+1 MACs for computation of update relation for weight. In case of signed regressor algorithm only one multiplication and accumulate operation is required to compute 'S.D(n)'. Whereas other two signed LMS algorithms does not require multiplication if we choose 'S' value a power of 2. In these cases multiplication becomes shift operation which is less complex in practical realizations. In SSA we Apply signum to With the purpose of coping up thru both complexity and issues in convergence deprived of restraining tradeoff, the corresponding signum based proportionate normalized and maximum proportionate normalized adaptive algorithms considered using LMS are Proportionate Normalized LMS (PNLMS), Data Clipped Proportionate Normalized LMS (DCPNLMS), Error Clipped Proportionate Normalized LMS (ECPNLMS), Sign Sign Proportionate Normalized LMS (SSPNLMS), Maximum Proportionate Normalized LMS (MPNLMS), Maximum Data Clipped Proportionate Normalized LMS (MDCPNLMS), Maximum Error Clipped Proportionate Normalized LMS (MECPNLMS), and Maximum Sign Sign Proportionate Normalized LMS (MSSPNLMS) algorithms. All these proposed algorithms provide lower complexity in computations for the reason that the presence of signum function in the technique and upright capability of filtering due to normalization term. These proportionate normalized and maximum proportionate normalized adaptive algorithms offers low computational complexity and good filtering capability compared to conventional LMS adaptive algorithm. The less complexity in computations of these adaptive algorithms leads to streamlined architecture aimed at lab on chip (LOC) or system on chip (SOC).

$$u(n+1) = u(n) + \frac{SG}{\varepsilon + G(\max(I(n)))^2} [I(n)]\, D(n) \quad (12)$$

The hybrid versions of MPNLMS with sign algorithms results in MDCPNLMS, MECPNLMS and MSSPNLMS algorithms.
The weight update equations of these maximum proportionate normalized signed algorithms are given as,

$$u(n+1) = u(n) + \frac{SG}{\varepsilon + G(\max(I(n)))^2} C[I(n)]\, D(n) \quad (13)$$

$$u(n+1) = u(n) + \frac{SG}{\varepsilon + G(\max(I(n)))^2} I(n)\, C[D(n)] \quad (14)$$

$$(n+1) = u(n) + \frac{SG}{\varepsilon + G(\max(I(n)))^2} C[I(n)]\, C[D(n)] \quad (15)$$

The proportionate normalized adaptive algorithms enjoy lower complexity in computations due to the presence of signum function in the algorithm. Finally projected techniques are successfully applied on real DNA sequences attained commencing National Center for biotechnological information (NCBI) genomic databank besides proved that they are more accurate for gene prediction.

both data and vector, and then we add 'S' to weight vector with addition with sign check (ASC) operation. Amongst all the techniques PNLMS algorithm is far complex; as this entails 2R+1 MACs and a division operation for implementation of update expression (8) for weight proceeding a processor based on DSP. In case of the maximum data clipped proportionate normalized LMS (MDCPNLMS) adaptive algorithm, computational complexity is less compared with other normalized algorithms with 1 MAC and 1 Division operations. It is evident that the shift and ASC operations require a lesser amount of circuitry of logic while related to MAC operations. However, by using a maximum normalization approach, we can minimize multiplications in the denominator from 'R' to '1'.

Related with further normalized technqiues, the MDCPNLMS technqiue needs lesser computations. For computing the variable step with low complexity in computations, the value of error produced in the principal iteration is adjusted and put in storage. Error part in next repetition is added to the previously stored value and squared. At that moment, result in order to be used in the next repetition is stored, and so on.

Curves for convergence for PNLMS with signum alternates are shown in Figure 2. Here, DCPNLMS is marginally lesser than PNLMS is evident. By reason of normalization in signum function is sensible to vector of data as well as normalization implemented above size of the step. Most important pro for DCPNLMS stands signum regressor job for multiplication operations are length liberated of the filter. Operations of sign regressor requires lone single multiplication. In this way, DCPNLMS performance is very near to PNLMS owing to dual normalizations made to lessen the multiplications by "R" amount, while, performance of SSPNLMS and ECPNLMS is improved compared to traditional LMS technique because of normalization, nevertheless lower than PNLMS and DCPNLMS technqiues as a result of error clipping that is accountable for updation of weight. Likewise, Figure 2 also displays the convergence curves for MPNLMS and the sign variants. Henceforth, normalization for size of the step accomplished using single portion in concert of MPNLMS along with alternates of signum function were poorer compared to PNLMS along with alternates of signum function. Residual facets for PNLMS seem effective than MPNLMS along with its alternates.

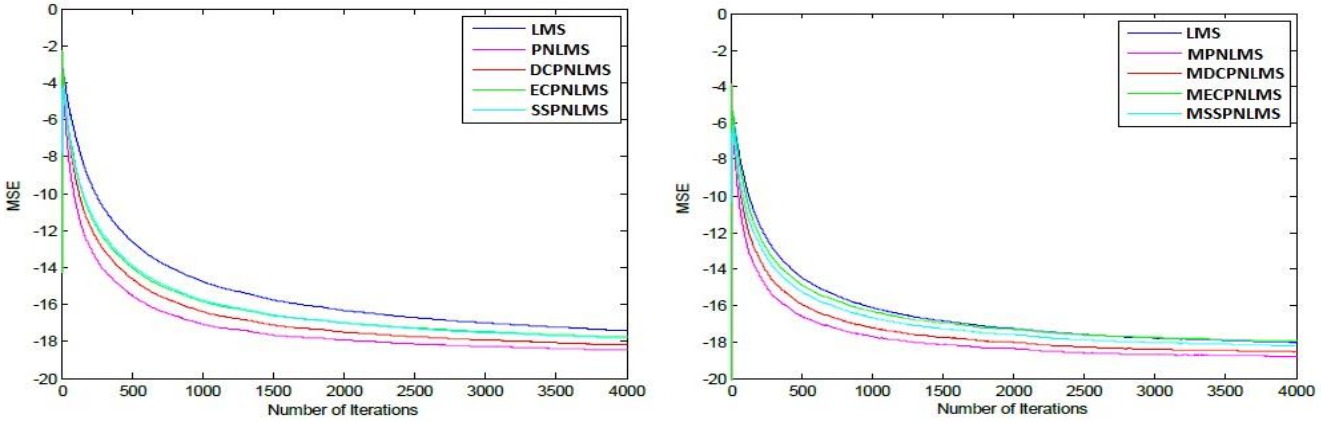


Fig. 2: Convergence characteristics of proportionate normalized LMS and its maximum proportionate normalized variants.

4. Results and Discussion

Here, the performances of several AEPs are presented and compared. The arrangement of proposed AEP is presented in Figure 1. Proportionate LMS with normalization and maximum proportionate normalized LMS algorithm with signum forms were accustomed for development of many AEP techniques. With the intent of assessment, an AEP based on LMS technique was presented in this work. In lieu of assessment, ten gene sequences were retrieved from NCBI gene databank [24]. Intended for uniformity of outcomes, performance of several algorithms is evaluated thru consideration of ten gene sequences for analysis. Explanation of databank deliberated was presented in Table 1.Performance measure is conceded expending factors like sensitivity (Sn), specificity (Sp) and precision (Pr). Expressions along with theory intended for the metrics were specified part of [18] [23]. Fallouts of locating the exon positions for gene signal 5 are presented part of Figure 3. Evaluation metrics Sn, Sp and Pr are restrained at threshold values from 0.4 to 0.9 with an interim of 0.05. At threshold value of 0.8 the exon identification is likely to be improved. Therefore, the vaues at threshold 0.8 are presented in Table 2.

Table 1: DNA sequences dataset from NCBI databank

Seq. no.	Accession no.	Sequence description
1	E15270.1	Osteoclast genesis inhibitory factor (OCIF) of human gene
2	X77471.1	Human tyrosine aminotransferase(tat) gene
3	AB035346.2	T-cell leukaemia/lymphoma 6(TCL6) Homo sapiens gene
4	AJ225085.1	Fanconi anaemia group A(FAA) Homo sapiens gene
5	AF009962	CC-chemokine receptor (CCR-5) Homo sapiens gene
6	X59065.1	Human acidic fibroblast growth factor(FGF) Homo sapiens gene
7	AJ223321.1	Transcriptional repressor(RP58) Homo sapiens gene
8	X92412.1	Titin (TTN) Homo sapiens gene
9	U01317.1	Sequence on chromosome 11 for Human beta globin
10	X51502.1	Gene for prolactin-inducible protein (GPIPI) for Homo sapiens

The steps for adaptive exon prediction are presented below:

a) DNA input sequences are selected from NCBI genome database. Using numerical technique of mapping, convert DNA sequence to numerical data. Provide attained digital data input given for structure of adaptive exon predictor as presented part of Figure 1.
b) A genomic sequence conforming base three periodicity specified to the adaptive exon predictor as a reference signal.
c) As depicted in Figure 1, a signal as feedback that is produced is used to apprise coefficients of filter.
d) The signal for feedback when turn out to be least, location of the exon region sequence is predicted precisely.
e) With help of power spectral density, location of the anticipated exon region is plotted. Performance metrics like Sn, Sp and Pr are calculated.

Figure 3 illustrates foreseen positions of the exon segments of sequence 5 applying several adaptive technqiues. Commencing these figures, it is vibrant that the LMS AEP has not predicted the exon segments correctly. This technique origins few uncertainties in exon prediction by detecting few intron segments. In Figure 3 (a) few adverse peaks are recognized at locations 1200th, 2300th and 3700th values of the sample. In unison, authentic position of required exon 4084-4268 was not forecasted. Prediction measures such as sensitivity, specificity and precision of PNLMS, DCPNLMS, ECPNLMS and SSPNLMS algorithms are observed as inferior than LMS adaptive algorithm where these are much better in case of maximum proportionate normalized algorithms. In the case of maximum proportionate normalized forms, the MPNLMS, MDCPNLMS, MECPNLMS and MSSPNLMS methods unerringly anticipated the true position of exon region at 4084-4268 thru decent PSD intensity are observed. The PSDs are shown in Figures 3 (f), (g), (h) and (i). Because of the normalization involved in these algorithms the tracking capability of these algorithms is better than LMS technique. Among these four algorithms MDCPNLMS is found to be better in connection with its complexity in computations along with characteristics of convergence. This algorithm needs less number of multiplications. The number of multiplications involved in this algorithm is independent of tap length of AEP. The convergence characteristics of MDCPNLMS are better than other normalized algorithms, though its performance measures are a bit inferior to PNLMS, MPNLMS, and MDCPNLMS algorithms. In the case of all proposed AEPs, the exon identification performance is superior to LMS and added normal signed variants. Consequently, depending on computational complexity, convergence characteristics, plots for exon identification, Sn, Sp and Pr calculations, it is found that MDCPNLMS based AEP is found to be the better candidate in practical applications.

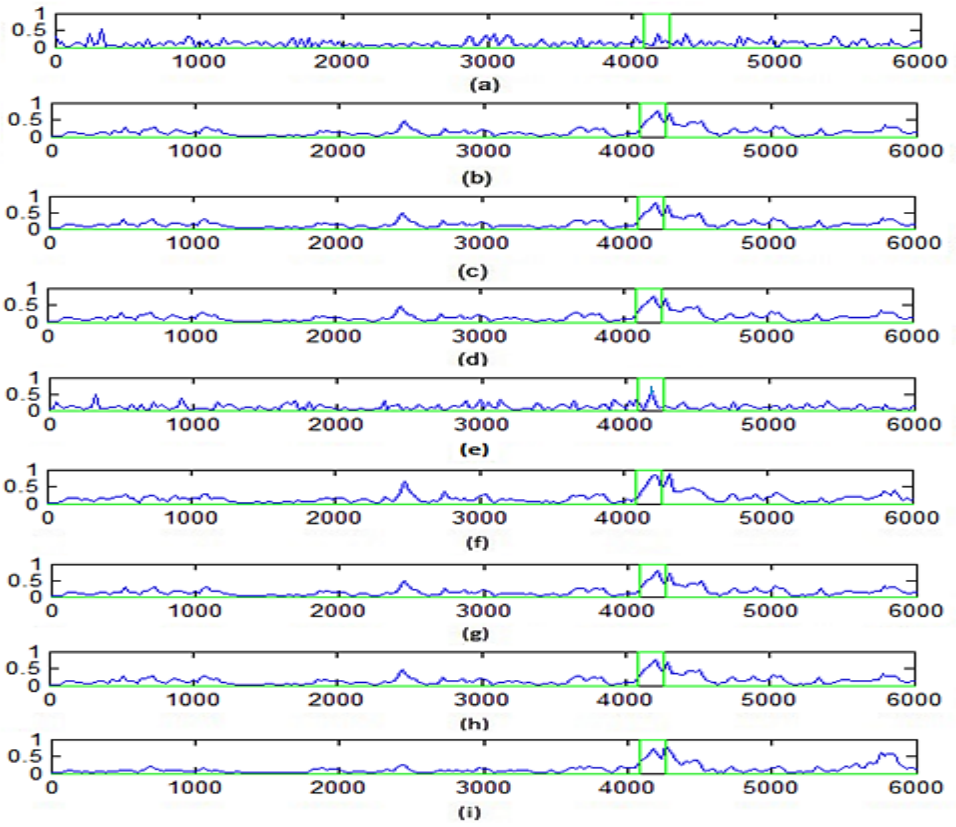


Fig. 3: Location of exons predicted using various proposed AEPs for genomic sequence with accession AF009962 (a). LMS based AEP, (b). PNLMS based AEP, (c). DCPNLMS based AEP, (d). ECPNLMS based AEP, (e). SSPNLMS based AEP, (f). MPNLMS based AEP (g).MDCPNLMS based AEP, (h). MECPNLMS based AEP and (i). MSSPNLMS based AEP

5. Conclusions

In this paper, the problem of identifying exons in a DNA sequence is illustrated. The concept of predicting the exact location of exons has several applications in current health care technology. At this point, we considered adaptive exon prediction techniques. To fulfill this we considered proportionate normalized and maximum proportionate normalized adaptive LMS algorithms to minimize the number of computations. In an attempt to further condense complexity in computations for projected implementations, we introduced the concept of proportionate normalization in addition to conventional LMS. To further minimize the complexity in computations, acclaimed PNLMS technqiue was united by signum based and maximum proportionate normalized signed algorithms. As a result eight novel algorithms that are amalgam in nature are considered and developed for predicting the exon segments in gene sequences. The hybrid variants are PNLMS, DCPNLMS, ECPNLMS, SSPNLMS, MPNLMS, MDCPNLMS, MECPNLMS and MSSPNLMS are considered for the current implementation. Different AEPs are developed and tested using these eight algorithms on real DNA sequences obtained from NCBI genome database. This is evident that MDCPNLMS AEP is improved in exon prediction applications, based on the convergence characteristics shown in Figure 2. This is also clear from the performance metrics charted as part of Table 3 along with the PSD for locating the exons is illustrated in Figure 3. Proposed AEPs exactly predicted the exon locations at 4084-4268 with good intensity as shown in PSD plot. The proposed MDCPNLMS based AEP based realization provides superior performance in terms of computational complexity based on performance measures with Sensitivity 0.7346, Specificity 0.7483 and precision 0.7325 obtained for a genomic sequence 5 with accession AF009962 as shown in Table 3 at a threshold value of 0.8. Therefore, the proposed normalized based AEPs are apt for real-world gene applications in developing the Nano devices, LOCs, and SOCs.

References

[1] L.W. Ning, H. Lin, H. Ding, J. Huang, N. Rao and F.B. Guo, Predicting bacterial essential genes using on sequence composition information, Genetics and Molecular Research, 13(2014), pp. 4564 - 4572.

[2] Min Li, Qi Li, Gamage Upeksha Ganegoda, JianXin Wang, Fang Xiang Wu, and Yi Pan, Prioritization of orphan disease-causing genes using topological feature and go similarity between proteins in interaction networks, SCIENCE CHINA Life Sciences, 57(2014), pp. 1064–1071.

[3] Dickerson JE, Zhu A, Robertson DL, and Hentges KE, Defining the role of essential genes in human disease, PloS One, 6(2011), e27368.

[4] Inbamalar T M, and Sivakumar R, Study of DNA Sequence Analysis Using DSP Techniques, Journal of Automation and Control Engineering, 1(2013), pp. 336–342.

[5] Cole S, Comparative myco bacterial genomics as a tool for drug target and antigen discovery, The European Respiratory Journal, 20(2002), pp. 78s–86s.

[6] S. Maji, D. Garg, Progress in gene prediction: principles and challenges, Current Bioinformatics, 8(2013), pp. 226–243.

[7] Hamidreza Saberkari, Mousa Shamsi, Hamed Heravi, and Mohammad Hossein Sedaaghi, A Novel Fast Algorithm for Exon Prediction in Eukaryotes Genes using Linear Predictive Coding Model and Goertzel Algorithm based on the Z-Curve, International Journal of Computer Appications, 67(2013), pp. 25–38.

[8] S. Maji and D. Garg, Progress in gene prediction, Current Bioinformatics, 8(2013), pp. 226– 243.

[9] Wazim Mohammed Ismail, Yuzhen Ye, and Haixu Tang, "Gene finding in metatranscriptomic sequences." BMC Bioinformatics, 15(2014), pp. 01–08.

[10] Mahin Ghorbani, Hamed Karimi, "Bioinformatics Approaches for Gene Finding." International Journal of Scientific Research in Science and Technology, 1(2015), pp. 12–15.

Copyright © 2018 Authors. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

[11] Gangchen Liu, Yihui Luan, "Identification of Protein Coding Regions in the Eukaryotic DNA Sequences based on Marple algorithm and Wavelet Packets Transform." Abstract and Appied Anaysis, 2014(2014), pp. 1-14.
[12] Yusuke Azuma, and Shuichi Onami, "Automatic Cell Identification in the Unique System of Invariant Embryogenesis in Caenorhabditis elegans." Biomedical Engineering Letters, 4(2014), pp. 328–337.
[13] BurraVenkataSrikanth, and Md Zia Ur Rahman, "Efficient ECG Signal Conditioning Techniques using Variable Step Size Least Mean Forth Algorithms." International Journal of Engineering and Technology, 8(2016), pp. 660-668.
[14] Srinivasareddy Putluri, and Md Zia Ur Rahman, "New Adaptive Exon Predictors For Identifying Protein Coding Regions In DNA Sequence." ARPN Journal of Engineering and Applied Sciences, 11(2016), pp. 13540-13549.
[15] Guangchen Liu and Yihui Luan, Identification of Protein Coding Regions in the Eukaryotic DNA Sequences based on Marple algorithm and Wavelet Packets Transform, Abstract and Applied Analysis, 2014(2014), pp. 01-14.
[16] Wagner K, and Doroslovacki M, Proportionate-type normalized least mean square algorithms, 59(2011), pp. 2410–2415.
[17] Md. Zia Ur Rahman, G.V.K.S. Karthik, S.Y. Fathima, A.L-Ekukaille, An efficient cardiac signal enhancement using time-frequency realization of leaky adaptive noise cancelers for remote health monitoring systems, Measurements, 46(2013), pp.3815-3835.
[18] Md. Zia Ur Rahman, Rafi Ahmed Shaik, D.V. Rama Koti Reddy, "Efficient and simplified Adaptive Noise Cancelers for ECG sensor Based Remote Health Monitoring", IEEE Sensors Journal, 91(2012), pp.566-573.
[19] Nagesh Mantravadi, S. V. A. V. Prasad, and Md. Zia Ur Rahman, "Artifact Removal in ECG signals using modified data normalization based signal enhancement units for healthcare monitoring systems." Journal of Theoretical and Applied Information Technology, 93(2011), pp. 225-239.
[20] Simon O. Haykin, Adaptive Filter Theory, 5th edition, Pearson Education Ltd., 2014.
[21] Md. Zia Ur Rahman, Rafi Ahamed Shaik, D. V. Rama Koti Reddy, "Efficient and Simplified Adaptive Noise Cancellers for ECG Sensor Based Remote Health Monitoring." IEEE Sensors Journal, 12(2012), pp. 566-573.
[22] Srinivasareddy Putluri, Md Zia Ur Rahman, "Simplified Adaptive Exon Predictors for extracting protein coding regions in genomic sequences." Journal of Theoretical and Applied Information Tecnhology, 93(2016), pp. 143 - 151.
[23] Paula S. R. Diniz, Adaptive Filtering, Algorithms and Practical Implementation, 3rd edition, Springer Publishers, 2014.
[24] National Center for Biotechnology Information, www.ncbi.nlm.nih.gov/.
[25]Thumbur Gowri, Injeti Sowmya, Md Zia Ur Rahman, D.V.R.K Reddy, "Adaptive Power Line Interference Removal from Cardiac Signals Using Leaky Based Normalized Higher Order Filtering Techniques", 2013 First International Conference on Artificial Intelligence, Modelling & Simulation, Malaysia, DOI 10.1109/AIMS.2013.54, 2013, pp. 294-298.
[26] Md. Zia Ur Rahman, S.R.Ahamed and D.V.R.K Reddy., "A Non-Linearities based Noise Canceller for Cardiac Signal Enhancement in Wireless Health Care Monitoring", IEEE Global Humanitarian Technology Conference, October 2012, USA.
[27] Md. Zia Ur Rahman, S.R.Ahamed and D.V.R.K Reddy ., "Cancellation of Artifacts in ECG Signals using Sign based Normalized Adaptive Filtering Technique", Proce. of 2009 IEEE Symposium on Industrial Electronics and Applications (ISIEA 2009), Malaysia, October 4-6,2009, pp. 442-445.
[28] Md. Zia Ur Rahman, S.R.Ahamed and D.V.R.K Reddy, "An Efficient noise Cancellation technique to remove noise from the ECG Signal using Normalized Signed Regressor LMS algorithm", Proce. of 2009 IEEE International Conference on Bioinformatics and Biomedicine (BIBM 2009), USA, Nov 1 – 4,2009, pp. 257-260.
[29] Md. Zia Ur Rahman, S.R.Ahamed and D.V.R.K Reddy., "Adaptive Noise Removal in the ECG using BLMS Algorithm," Proce. of 2nd IEEE International Conference on Adaptive Science & Technology (ICAST'09), Ghana, Africa, Dec 14-16,2009, pp. 380-383.
[30] Md. Zia Ur Rahman, S.R.Ahamed and D.V.R.K Reddy, "Noise Cancellation in ECG Signals using Normalized Sign-Sign LMS Algorithm", 9th IEEE International Symposium on Signal Processing and Information Technology(ISSPIT), UAE, Dec 14-17, 2009, pp. 288-292.
[31]Md. Zia Ur Rahman, S.R. Ahamed, D. V. R. K. Reddy, Ibrahim Khan, "Signed LMS based Adaptive Filtering to ECG Analysis : Noise Cancellation and Arrhythmia Detection", 2nd International Conference on Control, Instrumentation and Mechatronic Engineering (CIM2009), Mallacca, Malaysia, June 2-3, 2009.
[32]Md Nizamuddin Salman, P Trinatha Rao, Md Zia Ur Rahman, "Baseline Wander Removal in Cardiac Signals using Variable Step Size Adaptive Noise Cancellers", IEEE International Conference on Wireless Communications, Signal Processing and Networking, 23-25, March, 2016, Chennai, India. DOI: 978-1-4673-9338-6/16/$31.00_c 2016 IEEE, pp. 1529-1533.
[33]Md. Zia Ur, S.R.Ahamed and D.V.R.K Reddy, "Stationary and Non-Stationary noise removal from Cardiac Signals using a Constrained Stability Least Mean Square Algorithm", IEEE ICCSP 2011 , NIT Calicut , India, Feb, 10-12, 2011.
[34]Md. Zia Ur Rahman, V. Ajay Kumar, G V S Karthik, "A Low Complex adaptive algorithm for Antenna beam steering", IEEE 2011 International Conference on Signal Processing, Communications, Computing and Networking Technology (ICSCCN 2011), 2011, pp.317-321.
[35]Md. Zia Ur Rahman, S.R.Ahamed and D.V.R.K Reddy, "Baseline wander and Power line interference elimination from Cardiac Signals using Error Nonlinearity LMS algorithm ", IEEE ICSMB 2010 , IIT Kharagpur , India, Dec 16-18,2010.
[36]Md. Zia Ur Rahman, S.R.Ahamed and D.V.R.K Reddy, "Denoising ECG Signal using Transform Domain Adaptive Filtering Technique", IEEE INDICON 2009 , Ahmedaba , India, Dec 18-20,2009.
[37]Shafi Shahsavar Mirza, Md Zia Ur Rahman, "Efficient Adaptive Filtering Techniques for Thoracic Electrical Bio-Impedance Analysis in Health Care Systems", Journal of Medical Imaging and Heath Informatics, Vol.7, no-9, pp. 1126-1138, 2017.
[38]T. Gowri, P. Rajesh, Md.Zia Ur Rahman, D.V.R.K.Reddy, "Efficient ECG Signal Enhancement Techniques using Block Processed Noise Cancellers", Journal of Medical Imaging and Health Informatics, vol.6, no.3, pp.739-745, 2016.
[39]Md Zia Ur Rahman, Shafi Shahsavar Mirza,"Process Techniques For Human Thoracic Electrical Bio-Impedance Signal In Remote Healthcare Systems," IET Healthcare Technology Letters, DOI: 10.1049/Htl.2015.0061, pp. 1–5, 2016.
[40]Md.Zia Ur Rahman, S.R.Ahamed and D.V.R.K Reddy, "Noise Cancellation in ECG Signals using Computationally Simplified Adaptive Filtering Techniques: Application to Biotelemetry", Signal Processing: An International Journal, CSC Journals, ISSN 1985-2312, Vol. 3, Issue 5, pp. 1-12, 2009.
[41]Md. Nizamuddin Salman, P. Trinatha Rao, Md.Zia Ur Rahman, "Cardiac Signal Enhancement Using Normalised Variable Step Algorithm For Remote Healthcare Monitoring Systems," International Journal of Medical Engineering and Informatics, Inderscience Pub, Vol. 9, No. 2, 2017, pp. 145-161.
[42]Md. Zia Ur Rahman, Adaptive Noise Cancellers for Cardiac Signal Enhancement for IOT Based Health Care Systems, Journal of Theoretical and Applied Information Technology,Vol.95, no.10, 2017, pp.2206-2213.
[43]M. Nagesh, Md. Zia Ur Rahman, "A New ECG Signal Enhancement Strategy using Non-Negative Algorithms", International Journal of Control Theory and Applications Vol.10, no.35, 2017, pp.323-333.
[44]G V S Karthik, Md. Zia Ur Rahman, "ECG Signal Enhancement using Circular Leaky Adaptive Algorithm in an IOT Enabled Sensor System", International Journal of Control Theory and Applications,Vol.10, no.35, 2017, pp.271-282.
[45]Md. Salman, Md. Zia Ur Rahman, "Efficient and Low Complexity Noise Cancellers for Cardiac Signal Enhancement using Proportionate Adaptive Algorithms", Indian Journal Science and Technology,Vol.9, no-37, pp. 1-11, October 2016.
[46]M. Nagesh, Md. Zia Ur Rahman, "Efficient Noise Cancellers for ECG Signal Enhancement for Telecardiology Applications", Leonardo Electronic Journal of Practices and Technologies,Issue 29, 2016, pp.79-92.
[47]M. Nagesh, Md. Zia Ur Rahman, "Efficient Cardiac Signal Enhancement Techniques Based on Variable Step Size and Data Normalized Hybrid Signed Adaptive Algorithms", International Review on Computers and Software,Vol.11, no.10, 2016, pp.872-883.
[48]B. Srikanth, Md. Zia Ur Rahman, "Efficient ECG Signal Conditioning Techniques using Variable Step Size LMF Algorithms",

International Journal of Engineering and Technology,Vol. 8, No 2, pp.660-668, 2016.

[49]Asiya Sulthana, Md. Zia Ur Rahman, "Design and Implementation of Efficient Low Complexity Biomedical Artifact Canceller for Nano Devices", Leonardo Electronic Journal of Practices and Technologies,Issue 28, pp. 197-210, 2016.

[50]Md. Zia Ur Rahman, et. al., "Artifact Removal in ECG Signals using Modified Data Normalization Based Signal Enhancement Units for Health Care Monitoring Systems", Journal of Theoretical and Applied Information Technology,Vol.93, no.2, 2016, pp.540-5

Table 2. Performance measures of various AEPs with respect to Sn, Sp and Pr calculations.

Seq. No.	Parameter	LMS	PNLMS	DCPNLMS	ECPNLMS	SSPNLMS	MPNLMS	MDCPNLMS	MECPNLMS	MSSPNLMS
1	Sn	0.6286	0.7928	0.7772	0.7595	0.7381	0.7492	0.7407	0.7316	0.7202
	Sp	0.6435	0.7821	0.7636	0.7532	0.7265	0.7484	0.7323	0.7165	0.7112
	Pr	0.5922	0.7937	0.7623	0.7467	0.7388	0.7565	0.7452	0.7356	0.7223
2	Sn	0.6384	0.7824	0.7635	0.7569	0.7297	0.7491	0.7456	0.7232	0.7118
	Sp	0.6628	0.7932	0.7741	0.7575	0.7386	0.7485	0.7423	0.7376	0.7211
	Pr	0.5894	0.7836	0.7624	0.7515	0.7226	0.7463	0.7342	0.7157	0.7106
3	Sn	0.6457	0.7928	0.7782	0.7593	0.7381	0.7492	0.7417	0.7346	0.7206
	Sp	0.6587	0.7821	0.7636	0.7582	0.7265	0.7382	0.7123	0.7165	0.7112
	Pr	0.5934	0.7834	0.7723	0.7567	0.7388	0.7564	0.7432	0.7356	0.7223
4	Sn	0.6273	0.7945	0.7736	0.7535	0.7357	0.7438	0.7357	0.7274	0.7214
	Sp	0.6405	0.7824	0.7635	0.7589	0.7297	0.7591	0.7356	0.7232	0.7118
	Pr	0.5858	0.7937	0.7741	0.7575	0.7386	0.7485	0.7343	0.7276	0.7211
5	Sn	0.6481	0.7834	0.7624	0.7515	0.7326	0.7463	0.7346	0.7157	0.7106
	Sp	0.6518	0.7928	0.7712	0.7595	0.7361	0.7592	0.7483	0.7246	0.7202
	Pr	0.5904	0.7821	0.7636	0.7582	0.7365	0.7482	0.7325	0.7165	0.7112
6	Sn	0.6162	0.7945	0.7741	0.7536	0.7386	0.7525	0.7353	0.7276	0.7211
	Sp	0.6324	0.7834	0.7624	0.7415	0.7326	0.7563	0.7442	0.7257	0.7206
	Pr	0.5786	0.7928	0.7724	0.7584	0.7331	0.7592	0.7387	0.7146	0.7102
7	Sn	0.6193	0.7824	0.7624	0.7487	0.7292	0.7574	0.7476	0.7183	0.7121
	Sp	0.6529	0.7945	0.7736	0.7535	0.7357	0.7538	0.7357	0.7274	0.7214
	Pr	0.5896	0.7843	0.7626	0.7478	0.7262	0.7586	0.7374	0.7152	0.7097
8	Sn	0.6241	0.7982	0.7732	0.7582	0.7362	0.7587	0.7252	0.7195	0.7122
	Sp	0.6289	0.7836	0.7645	0.7454	0.7281	0.7564	0.7312	0.7281	0.7215
	Pr	0.5856	0.7954	0.7728	0.7572	0.7325	0.7567	0.7353	0.7257	0.7106
9	Sn	0.6268	0.7824	0.7635	0.7489	0.7297	0.7591	0.7356	0.7232	0.7118
	Sp	0.6452	0.7937	0.7741	0.7475	0.7286	0.7485	0.7343	0.7176	0.7101
	Pr	0.5814	0.7834	0.7624	0.7515	0.7226	0.7563	0.7442	0.7257	0.7206
10	Sn	0.6202	0.7923	0.7772	0.7527	0.7381	0.7592	0.7387	0.7146	0.7102
	Sp	0.6465	0.7821	0.7636	0.7482	0.7265	0.7482	0.7323	0.7265	0.7212
	Pr	0.5786	0.7934	0.7723	0.7567	0.7388	0.7564	0.7452	0.7256	0.7228