CELLULAR AUTOMATA: LINEAR PREDICTION OF NON-OVERLAPPING CODONS IN A GENOME EVOLUTION

Rama Naga Kiran Kumar. K 1, Ramesh Babu. I 2

1Research Scholar, Department of Computer Science & Engineering, Acharya Nagarjuna University, Guntur, Andhra Pradesh, India.
2Professor, Department of Computer Science & Engineering, Acharya Nagarjuna University, Guntur, Andhra Pradesh, India.

1 karasala_be@hotmail.com, 2 rinampudi@outlook.com

https://doi.org/10.26782/jmcms.2020.01.00017

Abstract

This research paper gives the idea of ‘non-overlapping n-ary codons’ is suggested as an innovative way to deal with the investigation of genome groupings in the system of analytical software engineering. Given a genome succession of length N, and one can have (N/n) non-overlapping n-ary codons with 0 or 1 or up to n-1 untouched nucleotides left in the arrangement. Fresh or unused nucleotides are not advised in the plan of genetic code.

Keywords: Non-Overlapping, Linear Prediction, n-ary Codons (n-codons), Genome Sequences

I. Introduction

Crick, Barnett, Watts-Tobin, and Brenner analytically demonstrated the idea of a DNA codon comprising of three DNA bases. The behavior of the codon was explained by Heinrich in 1961. Latterly, in 1997 Rajan Govindarajan E proposed the idea of non-overlapping n-ary codons. For instance, let us recognize the limited succession AGTCAGTCG of length 9. Presently one can have the accompanying non-overlapping codons of AGTCAGTCG: (i) AG, TC, AG, TC (Four non-covering 2-codons and an fresh nucleotide G); (ii) AGT, CAG, TCG (Three non-covering 3-codons); (iii) AGTC, AGTC (Two non-overlapping 4-codons and a fresh nucleotide G); and (v) AGTCA (One non-overlapping 5-codons and four fresh nucleotides G, T, C and G).

II. Arithmetical Narration of N-Ary Codons

For a n-ary codon set, one can have an exclusive numerical portrayal. In this paper the exclusive numerical portrayal of 2-ary, 3-ary, 4-ary, and 5-ary codons are defined. As indicated by this, one can have 16, 64, 256, 1024 numerical portrayals for 2-ary, 3-ary, 4-ary, 5-ary codons,
A. Non-overlapping Codons – Linear prediction

For contextual analysis, Brucella Suis 1330 genome succession of length 5806 is utilized here. The whole single strand of Brucella Suis 1330 genome succession is assessed as an arrangement of non-overlapping n-ary codons and linear prediction is completed by utilizing the octet calculation.

The results of this paper form the expectation of non-overlapping n-ary codons.

B. Algorithm: Octet Linear Prediction

Give us a chance to consider the succession \( x(n) \) of length 5806 comparing to the Brucella Suis 1330 genome. Expect that all these 5806 qualities are estimated estimations of a parameter related to a straight dynamical framework. These qualities are treated as velocities of the framework. Increasing speed is distinction between resulting speed esteems. Forecast of ninth speed esteem \( \vec{V}_9 \) soon after the 8-point arrangement \( x(n) \) is completed in two stages:

(i) \( \vec{V}_9 = a + 9b \) is the linear regression formula for prediction of 9th value \( \vec{V}_9 \) and
(ii) \( \vec{A}_8 = c + 8d \) is the linear regression formula for estimation of 8th acceleration value \( \vec{A}_8 \) and to get the predicted velocity for \( \vec{V}_9 \), subtract \( \vec{A}_8 \) from \( \vec{V}_9 \).

C. General Formula:

The difference between 9th estimated velocity and 8th estimated acceleration is the 9th predicted velocity

\[
9\text{th predicted velocity} = 9\text{th estimated velocity} - 8\text{th estimated acceleration}
\]

Then,

\[
a = \vec{V} - \frac{\sum x_i (\vec{x}_9 - \vec{V})}{\sum (\vec{x}_9 - \vec{V})^2}
\]

In other words,

\[
a = \vec{V} - 4.5 \left( \frac{\sum (\vec{x}_9 - 4.5)(\vec{x}_8 - \vec{V})}{\sum (\vec{x}_8 - 4.5)^2} \right)
\]

\[
\vec{V}_8 = \vec{V} - 4.5 \left( \frac{\sum x_i (\vec{x}_9 - 4.5)(\vec{x}_8 - \vec{V})}{\sum (\vec{x}_9 - 4.5)^2} \right) + 9 \left( \frac{\sum x_i (\vec{x}_9 - 4.5)(\vec{x}_8 - \vec{V})}{\sum (\vec{x}_9 - 4.5)^2} \right)
\]

\[
\vec{V}_9 = \vec{V} + 4.5 \left( \frac{\sum (\vec{x}_9 - 4.5)(\vec{x}_8 - \vec{V})}{\sum (\vec{x}_9 - 4.5)^2} \right) \text{ if } \vec{V}_9 \geq 0;
\]

\[
0 - \vec{V}_9 - \vec{V}_8 \text{ if } \vec{V}_9 < 0 \text{ (no further computation)}
\]
Instantly, the supreme estimated 9th value $V_9$ is given by:

$$V_9 = \lambda; \quad \text{when } \lambda \geq 0$$

$$= 0; \quad \text{when } \lambda < 0$$

D. Non-Overlapping 2-ary Codons: Linear Prediction

Below genome grouping is a non-overlapping 2-ary codons sequence of BrucellaSuis 1330 genome sequence. 2903 is the length of the below sequence. Fresh nucleotides are not available in this case.
Below are non-overlapping 2-ary codons numerical representation.
Below results are the calculated error sequence
The actual numerical values corresponding to non-overlapping 2-ary codons from 1 to 500 of the sequence, the predicted values along with the prediction error is shown in Fig. 1.

The actual numerical values corresponding to non-over 2-ary codons from 501 to 1000 of the sequence, the predicted values along with the prediction error is shown in Fig. 2.

Fig. 3 shows actual numerical values corresponding to non-over 2-ary codons from 1001 to 1500 of the sequence, the predicted values along with the prediction error.
The actual numerical values corresponding to non-over 2-ary codons from 1501 to 2000 of the sequence, the predicted values along with the prediction error is shown in Fig. 4.

The actual numerical values corresponding to non-over 2-ary codons from 2001 to 2500 of the sequence, the predicted values along with the prediction error shown in Fig. 5.

The actual numerical values corresponding to non-over 2-ary codons from 2501 to 3000 of the sequence, the predicted values along with the prediction error shown in Fig. 6.
E. Non-overlapping 3-ary codons: Linear Prediction

Below genome grouping is a non-overlapping 3-ary codons sequence of Brucella Suis 1330 genome sequence. 1935 is the length of the below sequence. One Fresh nucleotide is available in this case.

Prediction accuracy of the non-overlapping 2-ary codon is below:

\[
\text{Prediction Accuracy} = \frac{203}{2903} \times 100 = 6.992766145\% 
\]
Below is the numerical representation of the sequence of non-overlapping 3-ary codons.
Figs. 7 to 10 show genuine numerical qualities comparing to non-more than 3-ary codons from 1 to 1935 of the succession in steps of 500, the anticipated qualities alongside the forecast mistake.
Prediction accuracy of the non-overlapping 3-ary codon is below:

\[
\text{Prediction Accuracy} = \frac{38}{1935} \times 100 = 1.96382439\% 
\]

**F. Non-overlapping 4-ary codons: Linear Prediction**

Below genome grouping is a non-overlapping 4-ary codons sequence of Brucella Suis 1330 genome sequence. 1451 is the length of the below sequence. Two Fresh nucleotides are available in this case.
Below is the numerical representation of the sequence of non-over 4-codons.
The error sequence is evaluated, and the results are given below.

\[ \frac{1}{2} \sum_{i=1}^{n} (x_i - \hat{x}_i)^2 \]

where \( x_i \) is the actual value and \( \hat{x}_i \) is the predicted value.
Figs. 11 to 13 show genuine numerical qualities relating to more than 4-ary codons from 1 to 1451 of the grouping in steps of 500, the anticipated qualities alongside the expectation mistake.
Prediction accuracy of the non-overlapping 4-ary codon is below:

\[
\text{Prediction Accuracy} = \frac{9}{1451} \times 100 = 0.620261888 \%
\]

**G. Non-overlapping 5-ary codons: Linear Prediction**

Below genome grouping is a non-overlapping 5-ary codons sequence of Brucella Suis 1330 genome sequence. 1161 is the length of the below sequence. One Fresh nucleotide is available in this case.

Below is the numerical representation of the non-over 5-ary codons.

---

Rama Naga Kiran Kumar, K et al

234
Figs. 14 to 16 show genuine numerical quantities comparing to more than 5-ary codons from 1 to 1161 of the succession in steps of 500, the anticipated qualities alongside the prediction error.
III. Observation

Non-Overlapping codons Prediction accuracies

Table I gives the information about non-overlapping n-ary codons linear prediction.

<table>
<thead>
<tr>
<th>Codon Sequence</th>
<th>Prediction Accuracy of Non-Overlapping</th>
</tr>
</thead>
<tbody>
<tr>
<td>1-codon</td>
<td>28.71662365</td>
</tr>
<tr>
<td>2-codon</td>
<td>6.992766145</td>
</tr>
<tr>
<td>3-codon</td>
<td>1.963824349</td>
</tr>
<tr>
<td>4-codon</td>
<td>0.620261888</td>
</tr>
<tr>
<td>5-codon</td>
<td>0.086132644</td>
</tr>
</tbody>
</table>
References


