Paper Analysis–Using the ADAP Learning Algorithm to Forecast the Onset of Diabetes Mellitus

Ahmed Ibrahim, Youssef Hany, Mohamed Mobarak, and Mahmoud Tarek

•

Faculty of Engineering

October University for Modern Sciences and Arts

1. Problem Definition

The severity of the late diagnosis of diabetes mellitus increases with time, which urges for an early forecast, especially for a high-risk population such as the Pima Indians.

2. Proposed Solution

Testing the ability of an early neural network model–by paper's standard– ADAP to forecast the problem in question on a targeted group. Taking into account a crossover point for sensitivity and specificity of "0.76".

3. Methodology

The methodology is in fact divided in two connected parts, those parts are:

1 **Fixed Matrix**: A Matrix that contains eight entries for our inputs, having its rows subdivided corresponding to the range of values of the inputs. This plays a mapping role with cross-referencing the variable array.

2 **Variable Array**: Its the array that is generated by the algorithm, it keeps changing during the learning process while having the Fixed Matrix unchanged. It is of a single row and n columns that corresponds with the size of the Matrix, each cell in said column contains a value that corresponds to strength of connection–correlation–between the responder, and the association.

3 **The learning or the Forecasting of the Algorithm**:

   1 ADAP reads cases in for each entry in the variable array.

   2 The rows that match to the input variables for the read cases established by the ADAP is activated.

   3 ADAP counts the number of + signs in the first column that are in an activated rows. If that count is greater than the threshold, the cell in the variable array corresponding to the column is flagged. Otherwise it is not flagged.

   4 The next column is then compared and the algorithm iterates until all columns have been processed.

   5 The values of the flagged cells in the variable array are summed. That sum is the FORECAST for the case.

6 If ADAP is not in the learning mode the flags in all cells are erased and a new case is read in.
7 If ADAP is in the learning mode the actual state or value that ADAP had just tried to forecast is read in.
8 ADAP now makes appropriate internal corrections. It generates an error estimate, then this value is added to each of the flagged cells.
9 All flags are now removed and a new case is read in.

4. Parameters Definition

1. Number of times pregnant
2. Plasma Glucose Concentration at 2 Hours in an Oral
3. Diastolic Blood Pressure (mm Hg)
4. Triceps Skin Fold Thickness (mm)
5. 2-Hour Serum Insulin Uh/ml)
6. Body Mss Index (Weight in kg / (Height in in))
7. Diabetes Pedigree Function (shown below, mainly to relate the likeliness to develop diabetes based on genetic and age factors)
8. Age (years)

$$DPF = \frac{\sum_i K_i(88 - ADM_i) + 20}{\sum_j K_j(ALC_j - 14) + 50} \tag{1}$$

For Equ[1] the parameters are defined as follows:

i ranges over all relatives, who had developed diabetes by the subject's examination date;

j ranges over all relatives, who had NOT developed diabetes by the subject's examination date;

K_x is the percent of genes shared by the relatives.

ADM_i is the age in years of relatives, when diabetes was diagnosed.

ACL_j is the age in years of relatives, at the last non-diabetic examination.

The constants 88 and 14 represent, with rare exception, the maximum and minimum ages at which relatives of the subjects in this study developed Diabetes, while, 20 and 50 were chosen to modify the output of the DPF to that of what is expected in a clinical base.

INPUT VARIABLES	# OF CATEGORIES	CATEGORIES
Number of Times Pregnant	3	{0,1,2} {3,4,5,6} {>7}
2 Hr Glucose Tolerance	6	{0-89.1} {89.2-107.1} {blank†, 107.2-123.1} {123.2-143.1} {143.2-165.1} {>165.2}
Diastolic BP	4	{blank} {1-76.1} {76.2-98.1} {>98.2}
Triceps Skin Fold	4	{blank} {1-25} {26-32} {>33}
2 Hr Serum Insulin	5	{blank} {1-110} {111-150} {151-240} {>241}
Body Mass Index	5	{1-22.814} {22.815-26.84} {blank, 26.841-33.55} {33.551-35.563} {>35.564}
Diabetes Pedigree Function	5	{0-.244} {.245-.525} {.526-.805} {.806-1.11} {>1.11}
Age	5	{21-24} {25-30} {31-40} {41-55} {>55}

Figure 1: *Ranges of Parameters*

5. Flowchart

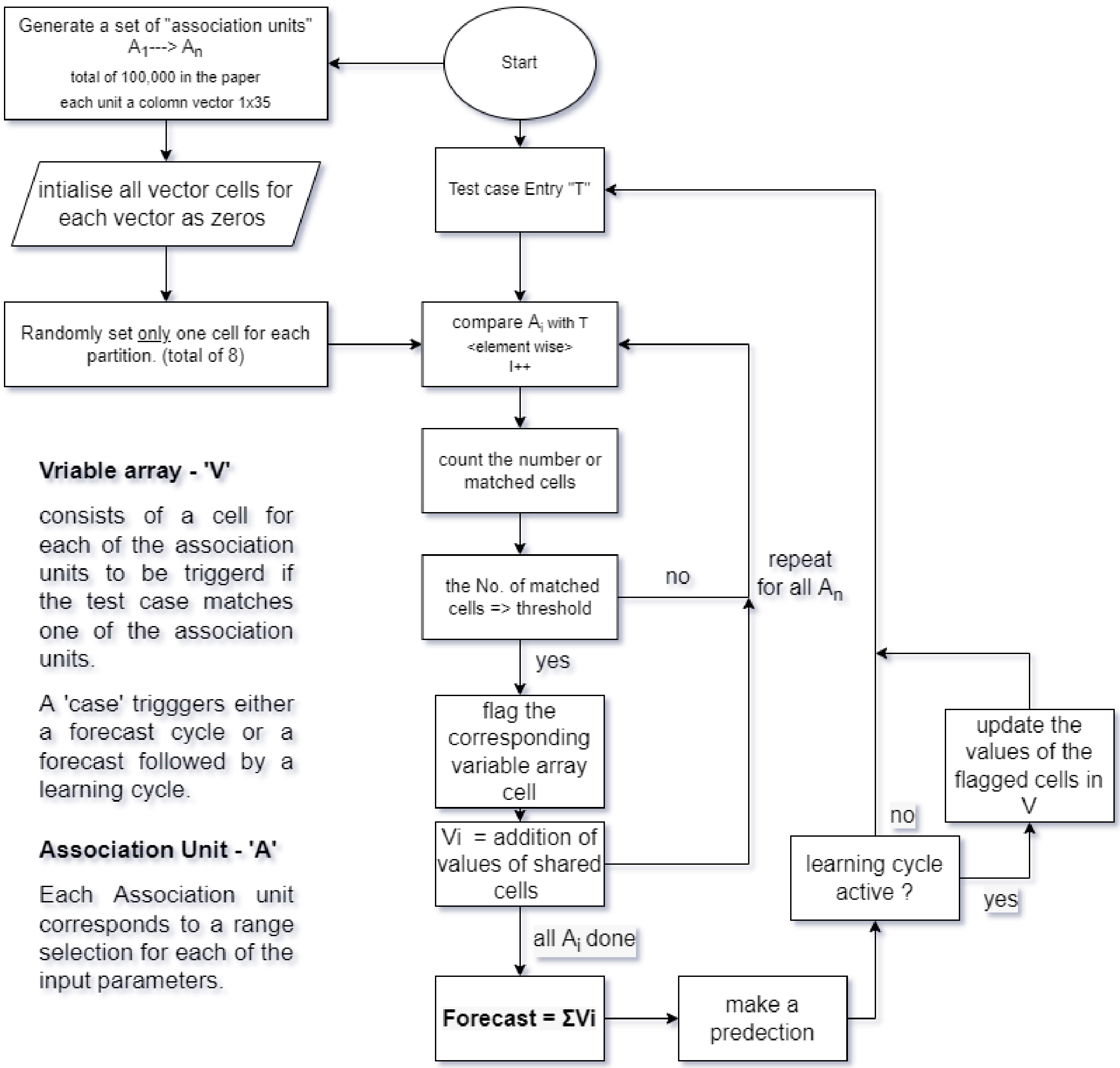

Figure 2: *Flowchart Design*

6. Success Criteria

Upon training the ADAP algorithm on a 576 cases it was pit into a non-learning mode –now what we call a best fit–, it produced what was referred to as the marginal value that being 0.448, in which it achieved the needed 0.76 crossover point and acted as a marginal difference.