**Flowchart for a novel SMS filtering approach**

**A. Perform a Pre-processing operation**

Start

Read in the SMS Message

Remove every unwanted character from the input SMS message

Convert the message to UTF-8 values of the characters in the text

Output the UTF-8 values of the SMS message

B

**B. Perform a 1D-TP transformation**

B

Get the UTF-8 values of the SMS message as an input

Set the initial value for the total number of utilized neighbours of characters, i.e. P

Set the initial value for the threshold parameter, i.e. Beta(B)

Arrange the input UTF-8 values of the SMS message into a cluster of P+1 characters, i.e. C

For each Clusters, Ck

Get the center or middle character of the Cluster, i.e. Pc

Get the list of Left neighbour characters of Pc by extract all the P/2 characters that comes before Pc in the given cluster as Pl

Get the list of Right neighbour characters of Pc by extract all the P/2 characters that comes after Pc in the given cluster as Pr

Merge the two lists (Pl and Pr) together into a new list as Q

For every ith character of the list Pk

Is

(Pc <= Qi + B AND Pc >= Qi - B)?

Is

(Pc > Qi + B)?

Set TPi = 1

Set TPi = 0

Set TPi = -1

Y

Y

N

N

C

D

E

C

Generate a Bit value (1) for every positive value of TP and a Bit value (0) for every negative value of TP, then add this to the left list BLPi

Generate a Bit value (0) for every positive value of TP and a Bit value (1) for every negative value of TP, then add this to the right list BRPi

Is

i = size of list Q?

Y

N

D

Generate the decimal value of the string of Bit values generated in the left list BLPk , and set this as the Upper feature value as UP = decimal(BLPk )

Generate the decimal value of the string of Bit values generated in the right list BRPi , and set this as the Lower feature value as LP = decimal(BRPk )

Output the two generated values (UP and LP) as the 1D-TP transformation of the input SMS message.

Is

All Cluster processed?

Y

N

E

F

**C. Perform a 1D-TP Histogram**

Generate a stream of bit values of size P (total number of utilized neighbours of characters)

F

Get the result of 1D-TP transformation as the input

For each of the 2 list UP and LP, generate a list of frequently occurred items into another 2 list HU and HL.

Generate the Histogram value (i.e. how often each of the patterns appear in its corresponding signal) of the new lists HU and HL

G

**D. Perform an Ant Colony Optimization**

Get the UTF-8 values of the SMS message as an input

G

Set initial value for the maximum possible iterations, as MaxIt

Set initial value for the total number of possible ants, as MaxAnt

Start a new Iteration, I

Select a new Ant

Set the initial value for the total number of utilized neighbours of characters, i.e. P

Set the initial value for the threshold parameter, i.e. Beta(B)

Arrange the input UTF-8 values of the SMS message into a cluster of P+1 characters, i.e. C

Move ant to randomly selected cluster of P+1 characters, C

Perform 1D-TP Transformation on the selected cluster C starting from its step 6(i.e. Step B6)

Perform 1D-TP histogram on the result obtained from 1D-TP transformation

Calculate the Cost

Update local pheromone trail

Is

Ant = MaxAnt?

Y

N

H

J

H

Get the best solutions from each of the result produce at every iteration

Update global pheromone

Is

is I = MaxIt?

Y

N

K

Output the best solution obtained.

J

**E. Perform a Classification Operation**

K

. Get the best solution obtained from K above (i.e. the optimal value of 1D-TP Histogram) as the input

Choose any of the classification methods (Multinomial Naïve Bayes, Bayesian network, random forest (RF), k nearest neighbors, radial basis feed forward neural network, relevance vector machine and artificial neural network)to classify the SMS message as ham or spam message.

Get the report of each classifier

Stop