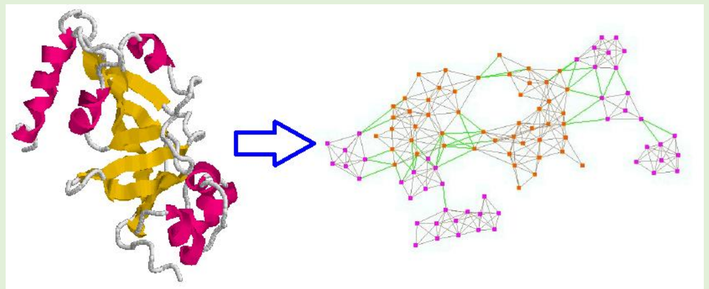
Protein Classification through Interaction Network Prediction using Multi-objective Bat Optimization

Md. Al-Imran and Md. Shiplu Hawlader

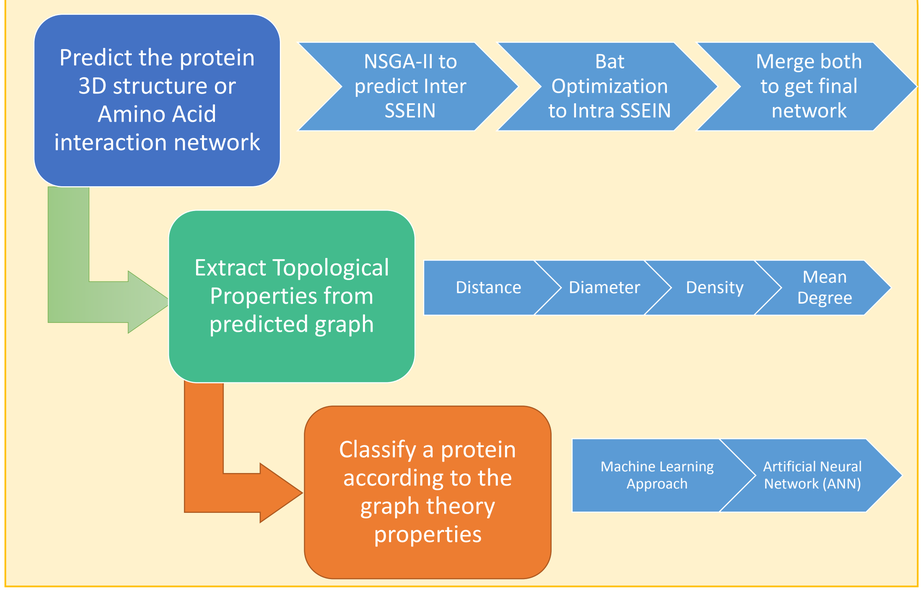
Department of Computer Science and Engineering, University of Dhaka

PROTEIN 3D STRUCTURE PREDICTION

Protein can be represented by amino acid interaction network. This is a graph whose vertices are the proteins amino acids and whose edges are the interactions between them. We can consider the protein structure prediction to actually prediction of amino acid interaction network.



CONTRIBUTIONS



OBJECTIVES

- ❑ Many of Protein's structure are difficult to discovering by the experiment.
- ❑ Discover the structure without experiment.
- ❑ Protein structure discovering experiments are too much expensive.
- ❑ Also can possible to the evaluation nature.
- ❑ To discover new drugs.
- ❑ Treat the predicted network as a graph.
- ❑ Classify a protein in computation way.

Observations

1. Distance between two Amino Acid.
   - Nearer is better
2. Torsion Angle.
   - Have to be within a threshold
3. Hydrophobicity
   - Same types make a bond.

Objective Functions

Distance = $\sqrt{((x_1 - x_2)^2 + (y_1 - y_2)^2 + (z_1 - z)^2)}$

Torsion Angle: $n_1 = A \times B$ and $n_2 = B \times C$

$X = n_1.x \times n_2.x + n_1.y \times n_2.y + n_1.z \times n_2.z$

$nor = \sqrt{n_2.x \times n_2.x + n_2.y \times n_2.y + n_2.z \times n_2.z}$

$y = n_1.x \times nor + n_1.y \times nor + n_1.z \times nor$

$\Phi = -\cos^{-1} x/y$

Hydrophobicity: If $p_1.h = 1$ (When Hydrophobic) and $p_1.h = 0$ (When Hydrophilic)

Than, Hydrophobicity = ($p_1.h$ XOR $p_2.h$)

Criteria for Bat Algorithm

Movement of virtual Bat:

$f_i = f_{min} + (f_{max} - f_{min})\beta$,Where $\beta \in [0,1]$

$v_i^t = v_i^{t-1} + (x_i^t - x^*) f_i$

$x_i^t = x^t{}_{i-1} + v_i^t$

Update of Loudness A & pulse rate r :

$A_i^{t+1} = \alpha A_i^t$

$r_i^{t+1} = r_i^{\gamma} [1 - e^{-\gamma t}]$,Where α ,γ are constants.

Algorithms to predicted the interaction network

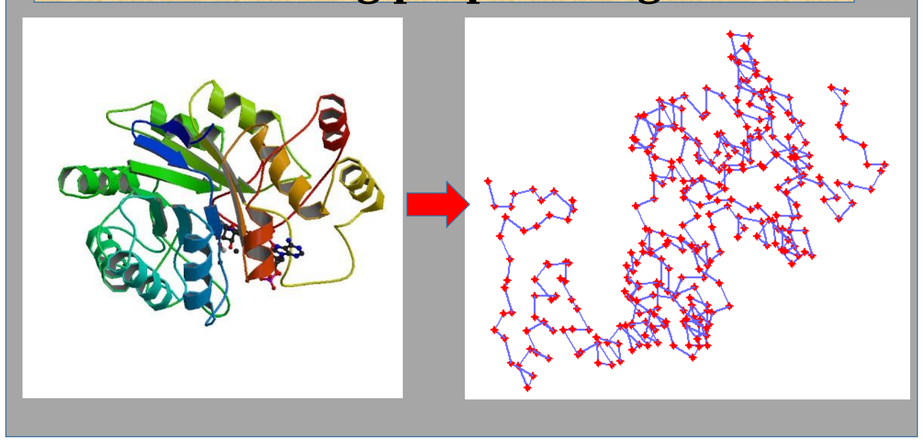
NSGA-II to predict inter SSE-IN

Data: PDB file of a protein
Result: The best predicted inter SSE amino acid interaction Network
Initialize No. of Objective M, No. of decision variable V
Initialize the population P
for $i \leftarrow 1$ *to* V **do**
  for $j \leftarrow 0$ *to* $No.of arrayindex$ **do**
    if $i = array[j]$ **then**
      lb = array[j], ub = array[j+1]
      Randomly chose a residue no. which not within lb and ub.
    end
  end
end
Calculate the initial fitness
Sort the population among it's Fitness
Calculate the initial Rank and Crowding Distance
for $i \leftarrow 1$ *to* T **do**
  Select Parent chromosome base on the Rank and Crowding distance value
  Make Genetic Operation Crossover and Mutation
  F(chromosome) = evaluate objective(chromosome)
  if $Fitness_{new} > Fitness_{old}$ **then**
    Update the population and Objective values
  end
  Sort the population among it's Fitness
  Calculate the new Rank and Crowding Distance
end

Bat Optimization to predict intra SSE-IN

Data: PDB file of a protein
Result: The best predicted intra SSE amino acid interaction Network
Initialize f_{min}, f_{max}, r_i,A_i
Initialize the population P
for $i \leftarrow 1$ *to* V **do**
  for $j \leftarrow 0$ *to* $No.of arrayindex$ **do**
    if $i = array[j]$ **then**
      lb = array[j], ub = array[j+1]
      Randomly chose a residue no. which within lb and ub.
    end
  end
end
Calculate the initial fitness
Calculate the initial Best solution
for $i \leftarrow 1$ *to* $no.of generation$ **do**
  Generate new solutions by adjusting frequency
  Updating velocities v_i and locations x_i
  if $rand > r_i$ **then**
    Select a solution among the best solutions
    Generate a local solution around the selected best solution
  end
  f(x) = evaluate objective()
  if $rand < A_i \& f(x_i) < f(x_*$ **then**
    Accept the new solutions Increase r_i and reduce A_i
  end
  Rank the bats and find the current best x^*
end

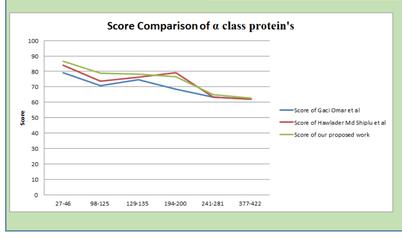
Predicted Amino Acid interaction network using proposed algorithm

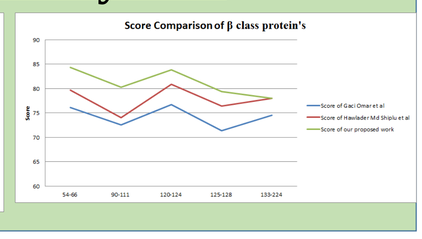


Results Analysis

By observing the resulted network, we can see that, some regions of this network have are dense and remain area are sparse. There also exists few edges between dense and sparse regions. It means that, this network predicted the intra SSE-IN and inter SSE-IN connections of this protein.

Score Analysis





Recognize Topological graph properties

Diameter D = max{d(u,v)}, where u,v ∈ V

Mean distance $d_G = \frac{n}{n(n-1)} \sum_{(u,v)\in V} d(u,v)$

Mean Degree $\bar{k_G} = \frac{1}{n} \sum_{(u,c)\in V} k_u = \frac{2m}{n}$

Density $\delta(G) = \frac{2m}{n(n-1)} \approx \frac{2m}{n^2}$

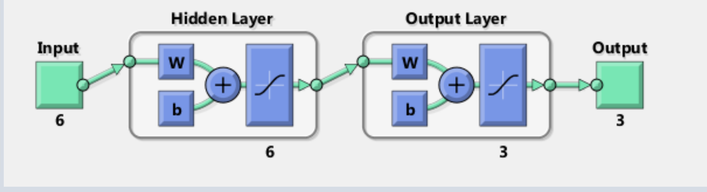
Result of Topological properties

Protein SCOP family	Diameter (D)	Mean distance (d_G)	Mean degree(k_G)	Density ($\delta(G)$)
Globin-like	13.61	4.12	25.25	0.074
TIM α / β Barrel	17.58	6.87	18.02	0.1831
Lysozyme	23.51	2.46	12.52	0.1038

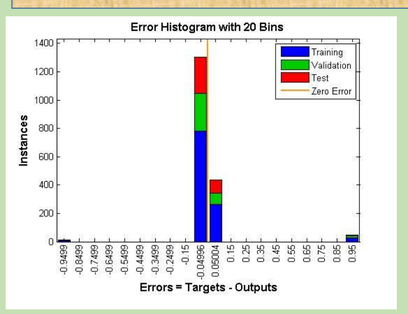
Classify the protein family

After observing the result of topological properties, it is clearly seen that Diameter and Mean degree of the network strongly depends on protein family, so we can able to make a classifier on these properties. We have used Artificial Neural Network to do the classification

Architecture of ANN



Result of the classification



The overall accuracy of our classification is 92%, So may claim that the result is good enough.

Conclusion

- From the performance analysis, we can claim that our framework provides better amino acid interaction network without providing any previous information in contrast to the previous related works [1] [2].
- Also from the analysis, of graph properties we can claim that, our framework provides some meaning-full properties of amino acid interaction network, which were the main features to the machine learning approach to classify a protein family in computational way
- After observing all the result, we can say that our proposed method is a improvement compare to previous related works in terms of result as well as breadth of the research

References

- Gaci Omar Balev Stefan. Ant colony approach to predict amino acid interaction networks. In World Congress on Nature & Biologically Inspired Computing, volume 1, pages 17251730, 2009. Systems and Security (NSysS) 2015 International.
- Hawlader Md Shiplu Tareeq Saifuddin Md. Multiobjective approach in predicting amino acid interaction network using ant colony optimization. International Journal of Artitcial Intelligence & Applications, 5(1):113, 2014..
- Deb K, Agrawal S, Pratap A, and Meyarivan T. A fast elitist non-dominated sorting genetic algorithm for multi-objective optimization: Nsga-ii. In Proc. of 6th international conference on parallel problem solving on nature, Paris,France, September 2000. Springer.
- Xin-She Yang. A new metaheuristic bat-inspired algorithm. In Nature inspired cooperative strategies for optimization (NICSO 2010), pages 6574. Springer, 2010.

Contacts : roton1785@gmail.com, shiplu@cse.du.ac.bd