

ევოლუციური ალგორითმის (GA) ფიტნეს ფუნქციის ოპტიმიზირება

დნმ-ს სეკვენირების

დავით გოდერძიშვილი

სიტყვა გასაღებები: გენეტიკური ალგორითმი, დნმ, ფიტნეს ფუნქცია, სეკვენირება.

ბიოლოგიური თანმიმდევრობებს შორის მსგავსებების აღმოჩენა, მოლეკულურ ბიოლოგიაში და გენეტიკაში წარმოადგენს ერთ-ერთ მნიშვნელოვან საკითხს. მისი მეშვეობით იქმნება ფილოგენეტიკური ხეები¹. ბიოლოგიურ თანმიმდევრობებია დნმ, რნმ და ცილის მოლეკულა.

Msa (multiple sequence alignment) მიზნად ისახავს თანმიმდევრობების განლაგებას იმგვარად რომ მოხდეს მათ შორის მსგავსი ერთეულების დამთხვევის მაქსიმიზირება. აღნიშნული ამოცანა წარმოადგენს NP- სრულ ამოცანას რაც გულისხმობს რომ ჯერ არ არსებობს მისი პოლინომიალურ დროში ამოხსნის მეთოდი [1,2]. თანამედროვე პროგრამები რომლებიც გამოიყენება ბიოლოგიური თანმიმდევრობების ანალიზისთვის (კრიმინალისტიკაში, ფილოგენეტიკა და სხვ.) ძირითადად ენდობიან ჰერონისტიკას, დინამიური პროგრამირების კონცეპტების გამოყენებით, რადგან დაიზოგოს კომპიუტერული რესურსები, თუმცა მათი შესრულების დროც საკმაოდ მაღალია და მოითხოვს ძლიერ გამომთვლელ მანქანებს. მაგალითისთვის ერთ-ერთი პოპულარული ალგორითმი T-Coffee-ის შესრულების სიჩქარის ზედა ზღვარი არის $O(N^2 L^2) + O(N^3 L) + O(N^3) + O(NL^2)$ სადაც N არის მიმდევრობების რაოდენობა ხოლო L მათი სიგრძე.

პრაქტიკულ სიტუაციებში ხშირად საჭიროა თანმიმდევრობის განსაზღვრა $N > 300$ –სთვის რამაც წარმოშვა ჰერონისტიკული მეთოდების უკეთ ათვისების საჭიროება. ერთ ერთი მათგანი არის გენეტიკური ალგორითმები (GA), რომელიც მიეკუთვნება ძიების ოპტიმიზაციის სტოქასტურ-ჰერონისტიკული მეთოდს. მისი მეშვეობით შესაძლებელია დამაკმაყოფილებელი, თუმცა არა აუცილებლად ზუსტი პასუხის მიღება, მრავალ მსგავსი ტიპის ამოცანაზე.

გენეტიკური ალგორითმის გამოყენებით ხდება ამოცანის სავარაუდო ამოხსნების (ინდივიდების) შემთხვევითად შერჩევა, შემდეგ ხდება მათი ცვლილება ევოლუციური პროცესების მიხედვით ესენია: **მუტაცია, კროსინგოვერი და სელექცია**. მიღებული ამონახსნებიდან გადაირჩევა საუკეთესოები რომელთაგანაც ხდება შემდეგი თაობის დაკომპლექტება. საუკეთესოს გადარჩევა ხდება ე.წ ფიტნეს ფუნქციის გამოყენებით.

ბიოლოგიური მიმდევრობების ანალიზისთვის შემოთავაზებულია ფიტნეს ფუნქციის რამდენიმე ვარიანტი იგი შეიძლება იყოს ისეთი მარტივი, როგორც მიმდევრობების მატრიცის თითოეულ სვეტში ელემენტების იგივეობის შემოწმება ან ისეთი რთული როგორც არის მდებარეობის გათვალისწინება თითოეული ელემენტისთვის.

აღნიშნული მოხსენება მიზანდასახავს განსხვავებული ფიტნეს ფუნქციის წარმოდგენას რომელიც ხელს შეუწყობს ანალიზის დაჩქარებას და მრავალი მიმდევრობის შემოწმებას საშუალო სიმძლავრის კომპიუტერებზე.

¹ ფილოგენეტიკური ხე აჩვენებს სახეობების ევოლუციურ „ნათესავურ“ ურთიერდამოკიდებულებას

ლიტერატურა

1. <https://www.techfak.uni-bielefeld.de/~stoye/dropbox/reportB-96-06.pdf>
2. <https://arxiv.org/ftp/arxiv/papers/0901/0901.2747.pdf>

Genetic algorithm for multiple sequence alignment

Key words: genetic algorithm, MSA, fitness function

Genetic algorithm is stochastic search technique, used in many area of science, in our case we show how genetic algorithm can be used in multiple sequence alignment (MSA), to improve time complexity. Genetic algorithm (GA) is very new in area of MSA and has already proven its advantage compare to traditional methods.

This approach consists of simple genetic algorithm with its basic functionality and new fitness function as evaluation technique of “individuals”.