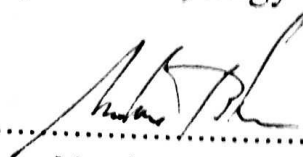


Sequencing of genomes of living organisms, that is discovering their linear structure (sequence of nucleotides) is now a basic way of collecting biological data. Such data can be analyzed and interpreted in several ways, but following the way to discover drugs one has to go into the higher level of research — prediction and modelling 3D structures of proteins and RNAs.

„Alulírott polgári- és büntetőjogi felelősségem tudatában kijelentem, hogy a másolat eredetivel mindenben megegyezik.”

2017. 06. 06.


.....
Madarasi Péter

that characterize specific local structural neighborhood observed around particular central atom, which is treated as a center of the descriptor. In principle, the descriptor is represented by a set of discontinuous fragments of a chain that are located around the atom. The important advantage of descriptors in comparison with continuous fragments of the chain is that descriptors take into consideration long range interactions that are observed in biomolecule structures. In our approach we developed the strategy to evaluate the similarity of descriptors using structural alignment methods. Previously, the problem of structural alignment of descriptors was posed and solved by some heuristic methods but in these cases, results of the alignments when descriptor A is matched to descriptor B, and vice versa, were often different. Our approach cures that weakness and assures symmetry in the procedure. Actually, the former papers presented the problem from the biological point of view being characterized by a multicriterial assessment function, and only biologically oriented results were given. Our strategies can be applied in the process of protein model quality assessment making the process of assessment more efficient and more accurate as well as can serve to develop structural libraries of descriptors that can represent fingerprints of particular biomolecules.

3 - Improved Algorithms for Matching Biological Graphs

Péter Madarasi, Alpár Jüttner

Subgraph isomorphism is a well-known NP-Complete problem, while its special case, the graph isomorphism problem is one of the few problems in NP neither known to be in P nor NP-Complete. Their appearance in many fields of application such as pattern analysis, computer vision questions and the analysis of chemical and biological systems has fostered the design of various algorithms for handling special graph structures.

The idea of using state space representation and checking some conditions in each state to prune the search tree has made the VF2 algorithm one of the state

of the art graph matching algorithms for more than a decade. Recently, biological questions of ever increasing importance have required more efficient, specialized algorithms.

This paper presents VF2++, a new algorithm based on the original VF2, which runs significantly faster on most test cases and performs especially well on special graph classes stemming from biological questions. VF2++ handles graphs of thousands of nodes in practically near linear time including preprocessing. Not only is it an improved version of VF2, but in fact, it is by far the fastest existing algorithm regarding biological graphs.

The reason for VF2++' superiority over VF2 is twofold. Firstly, taking into account the structure and the node labeling of the graph, VF2++ determines a state order in which most of the unfruitful branches of the search space can be pruned immediately. Secondly, introducing more efficient - nevertheless still easier to compute - cutting rules reduces the chance of going astray even further.

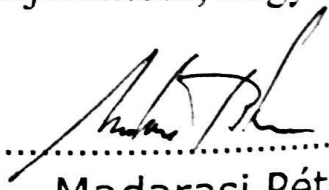
In addition to the usual subgraph isomorphism, specialized versions for induced subgraph isomorphism and for graph isomorphism are presented. VF2++ has gained a runtime improvement of one order of magnitude respecting induced subgraph isomorphism and a better asymptotical behaviour in the case of graph isomorphism problem.

After having provided the description of VF2++, in order to evaluate its effectiveness, an extensive comparison to the contemporary other algorithms is shown, using a wide range of inputs, including both real life biological and chemical datasets and standard randomly generated graph series.

The work was motivated and sponsored by QuantumBio Inc., and all the developed algorithms are available as the part of the open source LEMON graph and network optimization library (<http://lemon.cs.elte.hu>).

Alulírott polgári- és büntetőjogi felelősségem tudatában kijelentem, hogy a másolat edetivel mindenben megegyezik.”

2017. 06. 06.


.....
Madarasi Péter

Chair: *Alain Hertz*

1 - Combinatorial Analysis of the Pure Cycles of an FMS Consisting of Identical Work Centers