ON NEW SCALARIZATION TECHNIQUES IN MULTIOBJECTIVE OPTIMIZATION

Nikulin Y. V., Mäkelä M. M. , Wilppu O.

University of Turku, Turku, Finland, e-mail: yurnik@utu.fi

Most of the methods for multiobjective optimization utilize some scalarization technique where several goals of the original multiobjective problem are converted into a single-objective problem. One common scalarization technique is to use the achievement scalarizing functions. In our latest research [1, 2], we introduce a new family of two-slope parameterized achievement scalarizing functions for multiobjective optimization. This family generalizes both parametrized ASF and two-slope ASF. With these two-slope parameterized ASF, we can guarantee (weak) Pareto optimality of the solutions produced, and every (weakly) Pareto optimal solution can be obtained. The parameterization of this kind gives a systematic way to produce different solutions from the same preference information. With two weighting vectors depending on the achievability of the reference point, there is no need for any assumptions about the reference point. In addition to theory, we give graphical illustrations of two-slope parameterized ASF and analyze sparsity of the solutions produced in convex and nonconvex test problems.

References

1. Nikulin, Y. A new achievement scalarizing function based on parameterization in multiobjective optimization. / Y. Nikulin, K. Miettinen, M. Mäkelä // OR Spectrum. – 2012. – V. 34. P. 69–87.

2. Wilppu, O. New two-slope parameterized achievement scalarizing functions for nonlinear multiobjective optimization / O. Wilppu O., M. Mäkelä, Y. Nikulin // In: N. J. Daras, T. M. Rassias (eds.), Operations Research, Engineering, and Cyber Security, Springer Optimization and Its Applications. Springer. – 2017. – V. 113. P. 403–422.

RECONSTRUCTION OF DISEASE TRANSMISSIONS FROM VIRAL QUASISPECIES GENOMIC DATA

Skums P., Zelikovsky A., Dimitrova Z., Ramachandran S., Campo D., Bunimovich L., Khudyakov Y.

Georgia State University, Atlanta, GA, USA
Centers for Disease Control and Prevention, Atlanta, GA, USA
Georgia Institute of Technology, Atlanta, GA, USA
e-mail: skumsp@gmail.com

Genomic analysis is becoming a major tool for outbreak investigations. Existing computational frameworks for inference of transmission history from viral genomic data often do not consider intra-host diversity of pathogens and rely on additional epidemiological data, such as sampling times and exposure intervals. This impedes analysis of outbreaks of highly mutable viruses associated with chronic infections, such as HIV and HCV, whose transmissions are often carried out through minor intra-host variants, while the epidemiological information is unavailable or has a limited use.

The proposed framework QUENTIN (QUasispecies Evolution, Network-based Transmission INference)[1] addresses the above challenges by evolutionary analysis of intra-host viral populations sampled by deep sequencing and Bayesian inference using general properties of

social networks relevant to infection dissemination. It allows to infer transmission direction without additional case-specific epidemiological data, identify transmission clusters and reconstruct transmission history. QUENTIN was validated on data from 33 epidemiologically curated HCV outbreaks, yielding an accuracy of 87% for inference of transmission directions, 90% and 99.6% for detection of outbreak sources and transmission clusters, 78% and 98% for reconstruction of transmission links and ancestries. It was applied to investigate HCV transmissions within communities of hosts with high-risk behavior using the data collected during the investigation of several HIV/HCV outbreaks associated with drug abuse and commercial sex work. QUENTIN allowed to extract information on structures of transmission clusters, roles of transmission mode, co-infection and host's gender in the infection spread.

In conclusion, study of intra-host viral populations, evolutionary modelling and complex network analysis allow for accurate inference of disease transmissions. QUENTIN is most useful for investigation of extensively sampled outbreaks caused by RNA viruses. Its superior performance over consensus-based approaches indicates importance of quasispecies analysis for molecular surveillance and outbreak investigation.

References

1. QUENTIN: reconstruction of disease transmissions from viral quasispecies genomic data / P. Skums et al. // Bioinformatics. – 2017. – V. 34, N 1. – P. 163–170.

FAST ESTIMATION OF GENETIC RELATEDNESS BETWEEN MEMBERS OF HETEROGENEOUS POPULATIONS OF CLOSELY RELATED GENOMIC VARIANTS

Tsyvina V. A.

Georgia State University, Atlanta, e-mail: vyacheslav.tsivina@gmail.com

Many biological analysis tasks require extraction of families of genetically similar sequences from large datasets produced by Next-generation Sequencing (NGS). Such tasks include detection of viral transmissions by analysis of all genetically close pairs of sequences from viral datasets sampled from infected individuals or studying of evolution of viruses or immune repertoires by analysis of network of intra-host viral variants or antibody clonotypes formed by genetically close sequences. The most obvious naive algorithms to extract such sequence families are impractical in light of the massive size of modern NGS datasets. In this paper, we present fast and scalable k-mer-based framework to perform such sequence similarity queries efficiently, which specifically targets data produced by deep sequencing of heterogeneous populations such as viruses. The tool is freely available for download at **https://github.com/vyacheslav-tsivina/signature-sj**

Further we will use the following notation: $S=\{s_1,s_2,\ldots,s_L\}$ – a sequence over the alphabet $\{A,C,G,T\}$; k -mer – any subsequence of length k; k -segment – k -mer that starts at a position $1+ik$, $i=0,1,2,\ldots$; $K(S)$ – the set of all k -mers of the sequence S; $R(S)$ – the family of all k -segments of the sequence S (possibly with repetitions); $h(S,Q)$ – Hamming distance between two sequences S and Q; $l(S,Q)$ – edit distance (Levenshtein distance) between two sequences S and Q. We say that two sequences S and Q are related if $l(S,Q)\le t$ or , where t is a given threshold.