

3 - The Dynamic RORO Stowage Planning Problem

Alastair Main, Dario Pacino, Filipe Rodrigues

The shipping industry's greenhouse gas emission reduction has received significant focus over the past years. One of the research areas is that of stowage planning for RORO vessels. Efficient stowage plans are necessary to reduce the turnaround time for vessels in port. Reducing turnaround time results in prolonged sailing time, allowing vessels to reduce fuel consumption through slow steaming. When RORO vessels have calls at several ports, they handle cargo as an approximately FILO queue. Therefore, cargo can potentially become blocked when stowing cargo for later ports, behind cargo with an earlier discharge port.

Planning the cargo assignment onboard the vessels also requires considering the arrival time of cargo at the port. Recent research assumes that all freight is available for stowage when the RORO vessels arrive at the port. However, this is not always the case. The unique elements of scheduling and generation of loading/discharge paths are therefore of academic interest.

We propose a novel mathematical model with a weighted objective function that minimizes the relationship between the fuel consumption cost and the revenue gained from shipping cargo. The model schedules the cargo loading sequence to reduce time spent handling and re-handling cargo at each port. The problem is studied for a single deck layout for a vessel calling multiple ports. Results of the mathematical model and accompanying metaheuristic will be presented.

4 - Inventory Routing Problem with Multiple Time Windows and Time-based Cost Structure

Nooshin Heidari, Ahmad Hemmati

We propose a mathematical model and a math-heuristic algorithm to solve an Inventory Routing Problem with multiple time windows and time-based cost structure. The problem is to manage distribution of products that are produced and consumed with different rates within a many-to-many distribution structure. A heterogeneous fleet is used to transport products. Vehicles are allowed to visit each node during its multiple predefined time windows. The total number of visits for each node is limited to a given number during time horizon. The min/maximum stock level for each product at each node is predefined. The quantity of product to be handled at each node is limited. We defined different shifts during time horizon in which economical and environmental costs of transportation and driver wage cost are varied. The other operational costs are fixed and variable handling costs. Given the complexity of the problem, we develop an Adaptive Heuristic to determine routing variables which combines with a mixed integer programming formulation to specify the rest of variables. The algorithm starts with applying a number of heuristics to decrease penalty function. Once the algorithm reaches to the original feasible region, other heuristics defined for finding the best solution of the original problem. The performance of the algorithm is assessed through several small and large-sized problems. The quality and running time provide us with promising results compared to the one reported by Gurobi.

The first draft of the human reference genome was published more than two decades ago. Due to the technological limitations, it was full of gaps, especially in highly repetitive regions of centromeres and telomeres. Since that time, the reference genome was updated and corrected several times; until now, the newest version is GRCh38. DNA sequencing has become much cheaper in the last few years, and large sequencing projects are more popular. Many countries are sequencing large populations of individuals from their countries, which results in population genetic maps, with the most common variants that appear in the population. However, people noticed that the reference genome GRCh38, although well-annotated, is not sufficient to represent a given population. Several attempts were made in this area, and a few reference genomes were already published (e.g. Korea, China, Japan).

We present the results of the sequencing project called Genomic Map of Poland. The first polish reference genome is created de novo based on the trio: mother, father and child. In the pipeline, we used several technologies: short read, long high-quality reads (PacBio HIFI), artificial long reads (stLFR), long-distance contact reads (HiC), and ultralong reads (Nanopore). The result scaffolds, spanning the whole chromosome, were compared with the GRCh38 reference genome, showing the differences between the references.

2 - Processing and validation of population scale genomic data

Pawel Wojciechowski

Implementing a project where thousands of samples are sequenced is a demanding task. It requires high throughput sequencers, high-performance computing infrastructure and large data storage, as well as human resources and money. From the computer science point of view, especially when taking into account the whole genome sequencing it is also a challenge. Collecting and sequencing samples involve analysis of huge volumes of data. Processing a single human sample of standard size from raw data into germline variant calling takes around 30 hours on a CPU cluster node. In the case of the project Genomic Map of Poland, where genomes of around 5000 inhabitants will be sequenced, part of the work was outsourced, which caused new problems, e.g. with data transfer. Additionally, we had to introduce an additional stage of checking the consistency of the obtained results. Despite high expected overall quality of the data, some inaccuracies can be present, for example, samples may be swapped or files associated with a sample may have a wrong content. The data analysis we propose is for verification of correctness of such sequencing results provided on a massive scale.

3 - Improving the matching of genomic data with the crowd knowledge

Artur Laskowski

Understanding the human genome is a task that scientists have devoted much attention to over the last few decades. The enormous development of technology has reduced the total cost of reading the genome of a single human from a billion to a thousand dollars. The technology is not perfect yet; machines provide the genome divided into many overlapping sequences. We use ALGA, the fastest and the most reliable de novo assembly method. Despite the extraordinary speed of ALGA, it would take months to complete the task of assembling the genomes of thousands of people on a dedicated server. Given the problem with the amount of time and computational power required, we decided to propose a moderately faster approach. We want to perform the differential assembly (DA). The method uses the fact that human genomes are 99% alike. We assemble the next genome using common parts of a previously constructed assembly graph. The problematic fragments (structural variants) would be resolved separately. One of the challenging steps of DA is deciding which sequences of a new genome are connected to which parts of the previously assembled graph. We created the heuristic method using digraph matching, which performs decently. However, we decided to create an open contest on the optil.io platform for resolving this troublesome step. The problem is available on the optil.io platform, known for hosting optimization problems. We hope to use the knowledge of a crowd to improve our algorithm.

■ WC-23

Wednesday, 12:30-14:00 - Y307

CBBM 2

Stream: EWG CBBM, EURO working group on Computational Biology, Bioinformatics and Medicine
Invited session

Chair: *Marta Szachniuk*

Chair: *Maciej Antczak*

1 - Resolving the reference genome in the Genomic Map of Poland

Aleksandra Swiercz