Report 1 on Implementing Wine Paper

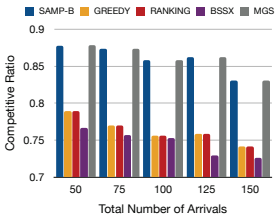
Soroush Vahidi

New Jersey Institute of Technology

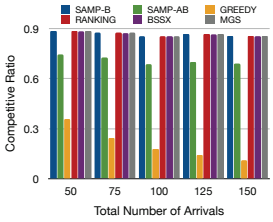
1 Summary of the Report

In this document, the approach and the results of implementing the greedy algorithm and ranking algorithm for both IFM and GFM and SAMP-B for only IFM. Moreover, it will be explained that why the offline part(LP) can be skipped for the greedy algorithm and the ranking algorithm. In most of the scenarios, answers found by Evan and this project are near; however, in some parts, especially about the greedy algorithm for GFM, the results have a difference. Moreover, it is shown that how the data is extracted from a very huge file that could not be read accurately by Microsoft Excel.

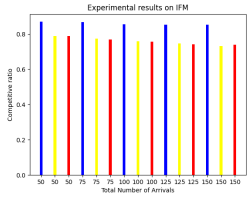
2 Tables



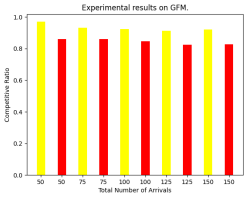
(a) Experimental results on IFM by Evan.



(b) Experimental results on GFM by Evan.



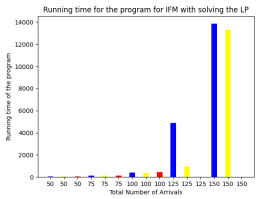
(c) Experimental results on IFM by Soroush.



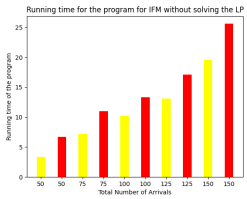
(d) Experimental results on GFM by Soroush.

Fig. 1: Experimental results of IFM and GFM on a real ride-hailing dataset in Chicago: The total number of arrivals T takes values from $\{50, 75, 100, 125, 150\}$ with $|I| = |J| = T$.

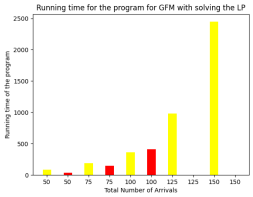
From figure 1, we can easily understand that the results of our implementation and Evan's are really similar except for the greedy algorithm for GFM. The author has checked his codes many times but has not found any bugs in his code. There is a little chance that the author has not imbibed the GFM meaning ,and maybe there is a bug in Evan's code. The answer for $T = 125$ and $T = 150$ for GFM in both of the algorithms are found by considering that the optimal answer of the LP is equal to 0.632; otherwise, the running time was very large.



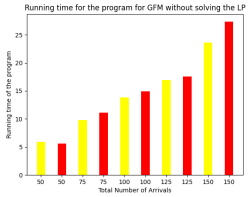
(a) Runtime for IFM with solving the LP.



(b) Runtime for IFM without solving the LP.



(c) Runtime for GFM with solving the LP.



(d) Runtime for GFM without solving the LP.

Fig. 2: Comparison of runtime for IFM and GFM with and without solving the LP.

3 Guess

According to the repeated number of tests and calculations, almost always the average of the optimum answer found by the LP is equal to $(e-1)/e \approx 0.632$; therefore, the author thinks with a very good precise, we can always consider the average answer of the LP equal to 0.632 and do not solve it; in order to decrease the running time of the program significantly. Please notice that we are talking about the average optimal answer for a large number of random test cases made using a specific dataset. Of course, it is possible to figure out test cases with an optimal value much less than 0.632.

4 Extracting the data

Working with the input file was a little hard because its volume was very large. The author wrote a program to extract useful data for a specific period of time. Some data in the original huge input file were invalid because they did not have some necessary fields like origin and destination. There were 7557 useful driver and rider pairs at all, while the Wine paper has talked about 11228 trips.

5 Additional Comments

The author also has implemented SAMP-B for GFM and some parts of MGS; however, he preferred to report them in the next report. Moreover, in this data set, for each passenger, there is only one group to ride him; therefore, the author cannot understand very good that how the different groups can compete in this scenario.