DNA Rearrangement: PrefixSort and Greedy Approaches

Your Name

2024-11-22

1   Objectives

Describe the objective(s) of the project and how these will be accomplished. You must give the necessary context to make the document self-contained, i.e., explain the problem or domain of application considered, the algorithm(s) that will be analyzed, which particular algorithmic issue(s) will be subject to scrutiny, etc.

2   Experimental Setup

Describe the configuration used in the experiments. This implies the following: (1) indicate what kind of experiments will be conducted (i.e., indicate in which way the algorithm will be run and what will be measured) and what will be the particular parameters that will be used in those experiments (i.e., their numerical values); (2) provide a description of the computational environment in which the experiments are run (see Table 1).

Table 1: Computational environment considered.

CPU	Write here your Processor specs, RAM
OS	Write here your Operating system name and version
Java	Write here your Java version

3   Empirical Results

A summary of the experimental results is provided in Tables 2 and 3 in the Appendix, along with the statistical fitting of the data to different growth models.

Describe the results, in particular Figure 1.

```
## Warning: package 'ggnewscale' was built under R version 4.4.2
```
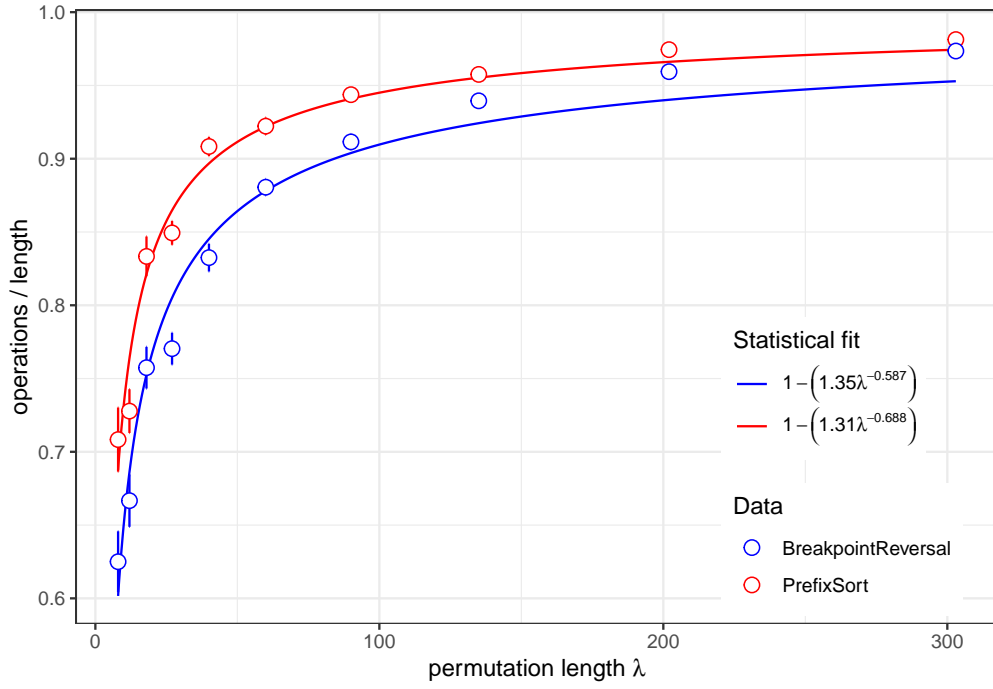
Figure 1: Reversals (normalized per number of elements) for increasing sequence length

4 Discussion

Provide your interpretation of the results: discuss whether the results match the theoretical predictions, whether some algorithm is better in practice than others, etc.

A Appendix

A.1 Data Summary

Table 2: Summary of the experimental results for PrefixSort. The mean and standard error are shown for each length.

length λ	operations	std. err.
8	0.708	2.19e-02
12	0.728	1.49e-02
18	0.833	1.36e-02
27	0.849	8.12e-03
40	0.908	6.49e-03
60	0.922	5.84e-03
90	0.944	4.12e-03
135	0.958	2.77e-03
202	0.974	1.90e-03
303	0.981	1.19e-03

Table 3: Summary of the experimental results for BreakpointReversal. The mean and standard error are shown for each length.

length λ	operations	std. err.
8	0.625	2.08e-02
12	0.667	1.79e-02
18	0.757	1.42e-02
27	0.770	1.09e-02
40	0.832	9.45e-03
60	0.881	5.60e-03
90	0.911	5.37e-03
135	0.940	3.78e-03
202	0.959	2.64e-03
303	0.973	1.55e-03

A.2 Model Fitting

```
## PrefixSort
## Formula: (1 - ops) ~ a * length^b
##
## Parameters:
##   Estimate Std. Error t value Pr(>|t|)
## a  1.30570    0.18836   6.932 0.000121 ***
## b -0.68790    0.05355 -12.845 1.27e-06 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.01719 on 8 degrees of freedom
##
## Number of iterations to convergence: 5
## Achieved convergence tolerance: 3.169e-06
```

```
##
## BreakpointReversal
## Formula: (1 - ops) ~ a * length^b
##
## Parameters:
##   Estimate Std. Error t value Pr(>|t|)
## a  1.35132    0.16127   8.379 3.12e-05 ***
## b -0.58740    0.04242 -13.846 7.16e-07 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.02061 on 8 degrees of freedom
##
## Number of iterations to convergence: 9
## Achieved convergence tolerance: 1.322e-06
```