Recompression of SLPs

Artur Jeż

Institute of Computer Science, University of Wrocław, Wrocław, Poland

Abstract

In this talk I will survey the recompression technique in case of SLPs. The technique is based on applying simple compression operations (replacement of pairs of two different letters by a new letter and replacement of maximal repetition of a letter by a new symbol) to strings represented by SLPs. To this end we modify the SLPs, so that performing such compression operations on SLPs is possible. For instance, when we want to replace ab in the string and SLP has a production $X \to aY$ and the string generated by Y is bw, then we alter the rule of Y so that it generates w and replace Y with bY in all rules. In this way the rule becomes $X \to abY$ and so ab can be replaced, similar operations are defined for the right sides of the nonterminals. As a result, we are interested mostly in the SLP representation rather than the string itself and its combinatorial properties. What we need to control, though, is the size of the SLP. With appropriate choices of substrings to be compressed it can be shown that it stays linear.

The proposed method turned out to be surprisingly efficient and applicable in various scenarios: for instance it can be used to test the equality of SLPs in time $\mathcal{O}(n \log N)$, where n is the size of the SLP and N the length of the generated string; on the other hand it can be used to approximate the smallest SLP for a given string, with the approximation ratio $\mathcal{O}(\log(n/g))$, where n is the length of the string and g the size of the smallest SLP for this string, matching the best known bounds.

1998 ACM Subject Classification F.2.2 Nonnumerical Algorithms and Problems

Keywords and phrases Straight Line Programs, smallest grammar problem, compression, processing compressed data, recompression

Digital Object Identifier 10.4230/LIPIcs.CPM.2017.2

Category Invited Talk

28th Annual Symposium on Combinatorial Pattern Matching (CPM 2017).
Editors: Juha Kärkkäinen, Jakub Radoszewski, and Wojciech Rytter; Article No. 2; pp. 2:1–2:1

Leibniz International Proceedings in Informatics
Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Dagstuhl Publishing, Germany