Two talks on variants of the Longest Filled Common Subsequence Problem

Alex Popa and Radu-Stefan Mincu

University of Bucharest

1. At CPM 2017, Castelli et al. define and study a new variant of the Longest Common Subsequence Problem, termed the Longest Filled Common Subsequence Problem (LFCS). For the LFCS problem, the input consists of two strings A and B and a multiset of characters $\mathcal{M}$. The goal is to insert the characters from $\mathcal{M}$ into the string B, thus obtaining a new string B^*, such that the Longest Common Subsequence (LCS) between A and B^* is maximized. Casteli et al. show that the problem is NP-hard and provide a 3/5-approximation algorithm for the problem.

In this talk we study the problem from the experimental point of view. We introduce, implement and test new heuristic algorithms and compare them with the approximation algorithm of Casteli et al. Moreover, we introduce an Integer Linear Program (ILP) model for the problem and we use the state of the art ILP solver, Gurobi, to obtain exact solution for moderate sized instances.

2. In Discrete Applied Mathematics 2010, Adi et al. introduce and study a variant of the well known Longest Common Subsequence problem, named *Repetition Free Longest Common Subsequence (RFLCS)*. In RFLCS the input consists of two strings A and B over an alphabet Σ and the goal is to find the longest common subsequence containing only distinct characters from Σ. Adi et al. prove that the problem is $\mathcal{APX}$-hard and show three approximation algorithms. Castelli et al. (Operations Research Letters 2013) propose a heuristic genetic algorithm and Blum and Blesa introduce metaheuristic algorithms (International Conference on Artificial Evolution 2013 and Evolutionary Computation in Combinatorial Optimization 2016).

In this talk we design and test several new heuristic algorithms for RFLCS. The first algorithm, uses dynamic programming and in our testing setup outperforms the algorithms of Adi et al.. The second heuristic algorithm improves upon the first and becomes comparable to the state-of-the-art al-

gorithms of Blum and Blesa. The third algorithm transforms the RFLCS instance into an instance of the Maximum Independent Set (MIS) problem with the same value of the optimum solution. Then, we apply known algorithms for the MIS problem. We also augment one of the approximation algorithms of Adi et al. and we prove that we achieve an approximation of factor $2\sqrt{\min\{|A|, |B|\}}$.

Finally, we introduce a new variant of the LCS problem, named *Multiset Restricted Common Subsequence (MRCS)*, that is a generalization of RFLCS. We present an exact polynomial time algorithm for MRCS for constant size alphabet. Additionally, we show that MRCS admits a $2\sqrt{\min\{|A|, |B|\}}$ approximation.