On the Longest Common Subsequence Problem

Anna Gorbenko

Department of Intelligent Systems and Robotics
Ural Federal University
620083 Ekaterinburg, Russia
gorbenko.ann@gmail.com

Vladimir Popov

Department of Intelligent Systems and Robotics
Ural Federal University
620083 Ekaterinburg, Russia
Vladimir.Popov@usu.ru

Abstract

In this paper we consider an approach to solve the longest common subsequence problem. This approach is based on constructing logical models for the problem.

Keywords: the longest common subsequence problem, logical model, satisfiability problem, **NP**-complete

1 Introduction

Algorithms on sequences of symbols are used in many very different applications, groups (see e.g. [1]), rings (see e.g. [2, 3]), semigroups (see e.g. [4, 5]), bioinformatics (see e.g. [6]), robotics (see e.g. [7, 8]), compression (see e.g. [9, 10]), etc. These algorithms now form a fundamental part of computer science. Algorithms on sequences of symbols received a lot of attention recently (see e.g. [11, 12, 13, 14]). In particular, computationally hard problems is extensively studied (see e.g. [15, 16, 17, 18]). One of the most important problems in analysis of sequences is the longest common subsequence (LCS) problem (see e.g. [18]). An approach to solve this problem is described in [18]. This approach is based on constructing a logical model for the longest common subsequence problem.

In papers [18, 19, 20, 21, 22, 23] the authors considered some algorithms to solve different logical models (see also [24, 25, 26, 27, 28, 29, 30]). In particular,

we have obtained an explicit reduction from LCS to PSAT in [18]. In this paper we consider explicit reductions from LCS to SAT and 3SAT.

In our reductions we use Boolean functions ξ and τ from [18]. Let

$$\xi' = \wedge_{\substack{1 \leq i[1] \leq m, 1 \leq i[2] \leq m, \\ i[1] \neq i[2], 1 \leq j[1] \leq |S_{i[1]}|, \\ 1 \leq j[2] \leq |S_{i[2]}|, 1 \leq p \leq k}} \wedge_{1 \leq l \leq |\Sigma|} ((\neg y[i[1], j[1], p] \vee \neg y[i[2], j[2], p]$$

$$\vee \neg x[i[1], j[1], l] \vee x[i[2], j[2], l]) \wedge$$

$$(\neg y[i[1], j[1], p] \vee \neg y[i[2], j[2], p] \vee x[i[1], j[1], l] \vee \neg x[i[2], j[2], l])).$$

Theorem 1. $\xi \Leftrightarrow \xi'$.

Proof. Note that $\alpha \rightarrow \beta \Leftrightarrow \neg \alpha \vee \beta$. Hence

$$(y[i[1], j[1], p] \wedge y[i[2], j[2], p]) \rightarrow x[i[1], j[1], l] = x[i[2], j[2], l] \qquad (1)$$

is equivalent to

$$\neg(y[i[1], j[1], p] \wedge y[i[2], j[2], p]) \vee x[i[1], j[1], l] = x[i[2], j[2], l]. \qquad (2)$$

Since $\neg(\alpha \wedge \beta) \Leftrightarrow \neg \alpha \vee \neg \beta$, in view of (2), we obtain that (1) is equivalent to

$$\neg y[i[1], j[1], p] \vee \neg y[i[2], j[2], p]) \vee x[i[1], j[1], l] = x[i[2], j[2], l]. \qquad (3)$$

Note that $\alpha = \beta \Leftrightarrow (\neg \alpha \vee \beta) \wedge (\alpha \vee \neg \beta)$. Hence (3) is equivalent to

$$\neg y[i[1], j[1], p] \vee \neg y[i[2], j[2], p]) \vee$$

$$((\neg x[i[1], j[1], l] \vee x[i[2], j[2], l]) \wedge (x[i[1], j[1], l] \vee \neg x[i[2], j[2], l])). \qquad (4)$$

In view of $\alpha \vee (\beta \wedge \gamma) \Leftrightarrow (\alpha \vee \beta) \wedge (\alpha \vee \gamma)$, it is easy to see that (4) is equivalent to

$$(\neg y[i[1], j[1], p] \vee \neg y[i[2], j[2], p] \vee \neg x[i[1], j[1], l] \vee x[i[2], j[2], l]) \wedge$$

$$(\neg y[i[1], j[1], p] \vee \neg y[i[2], j[2], p] \vee x[i[1], j[1], l] \vee \neg x[i[2], j[2], l]). \qquad (5)$$

Hence (1) is equivalent to (5). Therefore, $\xi \Leftrightarrow \xi'$. □

Theorem 1 gives us an explicit reduction from LCS to SAT.

If $|\Sigma| = 1$, then

$$\varphi'[i, j] = (x[i, j, 1] \vee z_1[1, j, 1] \vee z_1[i, j, 2]) \wedge (x[i, j, 1] \vee \neg z_1[1, j, 1] \vee z_1[i, j, 2]) \wedge$$

$$(x[i, j, 1] \vee z_1[1, j, 1] \vee \neg z_1[i, j, 2]) \wedge (x[i, j, 1] \vee \neg z_1[1, j, 1] \vee \neg z_1[i, j, 2]) \wedge$$

$$(\wedge_{1 \leq l[1] \leq |\Sigma|, 1 \leq l[2] \leq |\Sigma|, l[1] \neq l[2]} ((\neg x[i, j, l[1]] \vee \neg x[i, j, l[2]] \vee z_2[i, j, l[1], l[2]]) \wedge$$

$$(\neg x[i, j, l[1]] \vee \neg x[i, j, l[2]] \vee \neg z_2[i, j, l[1], l[2]]))).$$

If $|\Sigma| = 2$, then

$$\varphi'[i,j] = (x[i,j,1] \vee x[i,j,2] \vee z_3[i,j,1]) \wedge (x[i,j,1] \vee x[i,j,2] \vee \neg z_3[i,j,1]) \wedge$$

$$(\wedge_{1 \leq l[1] \leq |\Sigma|, 1 \leq l[2] \leq |\Sigma|, l[1] \neq l[2]}((\neg x[i,j,l[1]] \vee \neg x[i,j,l[2]] \vee z_4[i,j,l[1],l[2]]) \wedge$$

$$(\neg x[i,j,l[1]] \vee \neg x[i,j,l[2]] \vee \neg z_4[i,j,l[1],l[2]]))).$$

If $|\Sigma| = 3$, then

$$\varphi'[i,j] = (\vee_{1 \leq l \leq |\Sigma|} x[i,j,l]) \wedge$$

$$(\wedge_{1 \leq l[1] \leq |\Sigma|, 1 \leq l[2] \leq |\Sigma|, l[1] \neq l[2]}((\neg x[i,j,l[1]] \vee \neg x[i,j,l[2]] \vee z_5[i,j,l[1],l[2]]) \wedge$$

$$(\neg x[i,j,l[1]] \vee \neg x[i,j,l[2]] \vee \neg z_5[i,j,l[1],l[2]]))).$$

If $|\Sigma| = 4$, then

$$\varphi'[i,j] = (x[i,j,1] \vee x[i,j,2] \vee z_6[i,j,1]) \wedge (\neg z_6[i,j,1] \vee x[i,j,3] \vee x[i,j,4]) \wedge$$

$$(\wedge_{1 \leq l[1] \leq |\Sigma|, 1 \leq l[2] \leq |\Sigma|, l[1] \neq l[2]}((\neg x[i,j,l[1]] \vee \neg x[i,j,l[2]] \vee z_7[i,j,l[1],l[2]]) \wedge$$

$$(\neg x[i,j,l[1]] \vee \neg x[i,j,l[2]] \vee \neg z_7[i,j,l[1],l[2]]))).$$

If $|\Sigma| > 4$, then

$$\varphi'[i,j] = (x[i,j,1] \vee x[i,j,2] \vee z_8[i,j,1]) \wedge$$

$$(\wedge_{3 \leq l \leq |\Sigma|-2}(\neg z_8[i,j,l-2] \vee x[i,j,l] \vee z_8[i,j,l-1])) \wedge$$

$$(\neg z_8[i,j,|\Sigma|-3] \vee x[i,j,|\Sigma|-1] \vee x[i,j,|\Sigma|]) \wedge$$

$$(\wedge_{1 \leq l[1] \leq |\Sigma|, 1 \leq l[2] \leq |\Sigma|, l[1] \neq l[2]}((\neg x[i,j,l[1]] \vee \neg x[i,j,l[2]] \vee z_9[i,j,l[1],l[2]]) \wedge$$

$$(\neg x[i,j,l[1]] \vee \neg x[i,j,l[2]] \vee \neg z_9[i,j,l[1],l[2]]))).$$

Let $\varphi' = \wedge_{1 \leq i \leq m, 1 \leq j \leq |S_i|} \varphi'[i,j]$. Let

$$\psi'[1,i,j] = \wedge_{\substack{1 \leq l[1] \leq k \\ 1 \leq l[2] \leq k \\ l[1] \neq l[2]}}((\neg y[i,j,l[1]] \vee \neg y[i,j,l[2]] \vee z_{10}[i,j,l[1],l[2]]) \wedge$$

$$(\neg y[i,j,l[1]] \vee \neg y[i,j,l[2]] \vee \neg z_{10}[i,j,l[1],l[2]])),$$

$$\psi'[1] = \wedge_{\substack{1 \leq i \leq m \\ 1 \leq j \leq |S_i|}} \psi'[1,i,j],$$

$$\psi'[2,i,j,l] = \wedge_{\substack{1 \leq j[1] \leq |S_i| \\ j[1] \neq j}}((\neg y[i,j,l] \vee \neg y[i,j[1],l] \vee z_{11}[i,j,j[1],l]) \wedge$$

$$(\neg y[i,j,l] \vee \neg y[i,j[1],l] \vee \neg z_{11}[i,j,j[1],l])),$$

$$\psi'[2] = \wedge_{\substack{1 \le i \le m \\ 1 \le j \le |S_i| \\ 1 \le l \le k}} \psi'[2, i, j, l],$$

$$\psi'[3, i, j, l] = \wedge_{\substack{1 \le j[1] \le |S_i| \\ 1 \le l[1] \le k \\ j[1] < j \\ l[1] \ge l}} ((\neg y[i, j, l] \vee \neg y[i, j[1], l[1]] \vee z_{12}[i, j, j[1], l, l[1]]) \wedge$$

$$(\neg y[i, j, l] \vee \neg y[i, j[1], l[1]] \vee \neg z_{12}[i, j, j[1], l, l[1]])),$$

$$\psi'[3] = \wedge_{\substack{1 \le i \le m \\ 1 \le j \le |S_i| \\ 1 \le l \le k}} \psi'[3, i, j, l].$$

If $|S_i| = 1$, then

$$\psi'[4, i, l] = (y[i, 1, l] \vee z_{13}[i, 1, l] \vee z_{13}[i, 2, l]) \wedge (y[i, 1, l] \vee \neg z_{13}[i, 1, l] \vee z_{13}[i, 2, l]) \wedge$$

$$(y[i, 1, l] \vee z_{13}[i, 1, l] \vee \neg z_{13}[i, 2, l]) \wedge (y[i, 1, l] \vee \neg z_{13}[i, 1, l] \vee \neg z_{13}[i, 2, l]).$$

If $|S_i| = 2$, then

$$\psi'[4, i, l] = (y[i, 1, l] \vee y[i, 2, l] \vee z_{14}[i, 2, l]) \wedge (y[i, 1, l] \vee y[i, 2, l] \vee \neg z_{14}[i, 2, l]).$$

If $|S_i| = 3$, then $\psi'[4, i, l] = \psi[4, i, l]$. If $|S_i| = 4$, then

$$\psi'[4, i, l] = (y[i, 1, l] \vee y[i, 2, l] \vee z_{15}[i, 1, l]) \wedge (\neg z_{15}[i, 1, l] \wedge y[i, 3, l] \wedge y[i, 4, l]).$$

If $|S_i| > 4$, then

$$\psi'[4, i, l] = (y[i, 1, l] \vee y[i, 2, l] \vee z_{16}[i, 1, l]) \wedge$$

$$(\wedge_{3 \le j \le |S_i| - 2}(\neg z_{16}[i, j - 2, l] \vee y[i, j, l] \vee z_{16}[i, j - 1, l])) \wedge$$

$$(\neg z_{16}[i, |S_i| - 3, l] \vee y[i, |S_i| - 1, l] \vee y[i, |S_i|, l]).$$

Let $\psi'[4] = \wedge_{1 \le i \le m, 1 \le l \le k} \psi'[4, i, l]$. Let

$$\xi'' = \wedge_{\substack{1 \le i[1] \le m, 1 \le i[2] \le m, \\ i[1] \ne i[2], 1 \le j[1] \le |S_{i[1]}|, \\ 1 \le j[2] \le |S_{i[2]}|, 1 \le p \le k}} \wedge_{1 \le l \le |\Sigma|} ((\neg y[i[1], j[1], p] \vee \neg y[i[2], j[2], p] \vee$$

$$z_{17}[i[1], j[1], i[2], j[2], p, l]) \wedge$$

$$(\neg z_{17}[i[1], j[1], i[2], j[2], p, l] \vee \neg x[i[1], j[1], l] \vee x[i[2], j[2], l]) \wedge$$

$$(\neg y[i[1], j[1], p] \vee \neg y[i[2], j[2], p] \vee z_{18}[i[1], j[1], i[2], j[2], p, l]) \wedge$$

$$(\neg z_{18}[i[1], j[1], i[2], j[2], p, l] \vee x[i[1], j[1], l] \vee \neg x[i[2], j[2], l])).$$

Let $\tau' = \varphi' \wedge (\wedge_{1 \le q \le 4} \psi'[q]) \wedge \xi''$.

Theorem 2. $\tau \Leftrightarrow \tau'$.

Proof. Since

$$\alpha \vee \beta \Leftrightarrow (\alpha \vee z) \wedge (\beta \vee \neg z) \tag{6}$$

where z is a new variable, it is easy to see that $\psi[4] \Leftrightarrow \psi'[4]$ and $\xi' \Leftrightarrow \xi''$. Note that

$$\alpha \Leftrightarrow (\alpha \vee z) \wedge (\alpha \vee \neg z). \tag{7}$$

Hence $\psi[1] \Leftrightarrow \psi'[1]$, $\psi[2] \Leftrightarrow \psi'[2]$, and $\psi[3] \Leftrightarrow \psi'[3]$. In view of (6) and (7), it is easy to check that $\varphi \Leftrightarrow \varphi'$. Therefore, $\tau \Leftrightarrow \tau'$. $\square$

Theorem 2 gives us an explicit reduction from LCS to 3SAT.

We have obtained explicit reductions from LCS to some variants of satisfiability, SAT and 3SAT.

There is a well known site on which solvers for SAT are posted [31]. They are divided into two main classes, stochastic local search algorithms and improved exhaustive search algorithms. All solvers allow the conventional format for recording DIMACS boolean function in conjunctive normal form and solve the corresponding problem [31]. In addition to the solvers the site also represented a large set of test problems in the format of DIMACS. This set includes a randomly generated problems of 3SAT.

We have created a generator of natural instances for LCS. Also, we have used test problems from [31].

We have used algorithms from [31]. Also, we have designed our own genetic algorithm for SAT which based on algorithms from [31].

We have used heterogeneous cluster based on three clusters (Cluster USU, Linux, 8 calculation nodes, Intel Pentium IV 2.40GHz processors; umt, Linux, 256 calculation nodes, Xeon 3.00GHz processors; um64, Linux, 124 calculation nodes, AMD Opteron 2.6GHz bi-processors) [32].

Each test was runned on a cluster of at least 100 nodes. The maximum solution time was 15 hours. The average time to find a solution was 18.7 minutes. The best time was 29 seconds.

References

[1] Yu. M. Vazhenin and V. Yu. Popov, Decidability boundaries for some classes of nilpotent and solvable groups, *Algebra and Logic*, 39 (2000), 73-77.

[2] V. Yu. Popov, A ring variety without an independent basis, *Mathematical Notes*, 69 (2001), 657-673.

[3] V. Yu. Popov, On equational theories of varieties of anticommutative rings, *Mathematical Notes*, 65 (1999), 188-201.

[4] V. Popov, The Word Problem for Relatively Free Semigroups, *Semigroup Forum*, 72 (2006), 1-14.

[5] V. Yu. Popov, The property of having an independent basis in semigroup varieties, *Algebra and Logic*, 44 (2005), 46-54.

[6] V. Popov, The approximate period problem for DNA alphabet, *Theoretical Computer Science*, 304 (2003), 443-447.

[7] A. Gorbenko, A. Lutov, M. Mornev, and V. Popov, Algebras of Stepping Motor Programs, *Applied Mathematical Sciences*, 5 (2011), 1679-1692.

[8] A. Gorbenko and V. Popov, A Real-World Experiments Setup for Investigations of the Problem of Visual Landmarks Selection for Mobile Robots, *Applied Mathematical Sciences*, 6 (2012), 4767-4771.

[9] V. Yu. Popov, NP-Completeness of the Dictionary Generation Problem, *Siberian Advances in Mathematics*, 16 (2006), 115-127.

[10] A. Gorbenko and V. Popov, On Face Detection from Compressed Video Streams, *Applied Mathematical Sciences*, 6 (2012), 4763-4766.

[11] V. Popov, Sorting by prefix reversals, *IAENG International Journal of Applied Mathematics*, 40 (2010), 24-27.

[12] A. Gorbenko, V. Popov, and A. Sheka, Robot Self-Awareness: Temporal Relation Based Data Mining, *Engineering Letters*, 19 (2011), 169-178.

[13] A. Gorbenko, V. Popov, and A. Sheka, Robot Self-Awareness: Exploration of Internal States, *Applied Mathematical Sciences*, 6 (2012), 675-688.

[14] A. Gorbenko and V. Popov, Robot Self-Awareness: Occam's Razor for Fluents, *International Journal of Mathematical Analysis*, 6 (2012), 1453-1455.

[15] V. Yu. Popov, Computational complexity of problems related to DNA sequencing by bybridization, *Doklady Mathematics*, 72 (2005), 642-644.

[16] V. Popov, Multiple genome rearrangement by swaps and by element duplications, *Theoretical Computer Science*, 385 (2007), 115-126.

[17] V. Popov, The Approximate Period Problem, *IAENG International Journal of Computer Science*, 36 (2009), 268-274.

[18] A. Gorbenko and V. Popov, The Longest Common Subsequence Problem, *Advanced Studies in Biology*, 4 (2012), 373-380.

[19] A. Gorbenko, M. Mornev, and V. Popov, Planning a Typical Working Day for Indoor Service Robots, *IAENG International Journal of Computer Science*, 38 (2011), 176-182.

[20] A. Gorbenko and V. Popov, Programming for Modular Reconfigurable Robots, *Programming and Computer Software*, 38 (2012), 13-23.

[21] A. Gorbenko and V. Popov, On the Optimal Reconfiguration Planning for Modular Self-Reconfigurable DNA Nanomechanical Robots, *Advanced Studies in Biology*, 4 (2012), 95-101.

[22] A. Gorbenko and V. Popov, The set of parameterized k-covers problem, *Theoretical Computer Science*, 423 (2012), 19-24.

[23] A. Gorbenko, V. Popov, and A. Sheka, Localization on Discrete Grid Graphs, *Proceedings of the CICA 2011*, (2012), 971-978.

[24] A. Gorbenko and V. Popov, The Longest Common Parameterized Subsequence Problem, *Applied Mathematical Sciences*, 6 (2012), 2851-2855.

[25] A. Gorbenko and V. Popov, The Binary Paint Shop Problem, *Applied Mathematical Sciences*, 6 (2012), 4733-4735.

[26] A. Gorbenko and V. Popov, Element Duplication Centre Problem and Railroad Tracks Recognition, *Advanced Studies in Biology*, 4 (2012), 381-384.

[27] A. Gorbenko and V. Popov, On the Problem of Placement of Visual Landmarks, *Applied Mathematical Sciences*, 6 (2012), 689-696.

[28] A. Gorbenko and V. Popov, The Problem of Selection of a Minimal Set of Visual Landmarks, *Applied Mathematical Sciences*, 6 (2012), 4729-4732.

[29] A. Gorbenko, M. Mornev, V. Popov, and A. Sheka, The problem of sensor placement for triangulation-based localisation, *International Journal of Automation and Control*, 5 (2011), 245-253.

[30] A. Gorbenko, M. Mornev, V. Popov, and A. Sheka, The Problem of Sensor Placement, *Advanced Studies in Theoretical Physics*, 6 (2012), 965-967.

[31] URL: http://people.cs.ubc.ca/∼hoos/SATLIB/

[32] URL: http://parallel.uran.ru/mvc_now/hardware/supercomp.htm