Applied Mathematical Sciences, Vol. 7, 2013, no. 13, 645 - 650

Longest Common Parameterized Subsequences with Fixed Common Substring

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 the problem of the longest common parameterized subsequence with fixed common substring (STR-IC-LCPS). In particular, we show that STR-IC-LCPS is **NP**-complete. We describe an approach to solve STR-IC-LCPS. This approach is based on an explicit reduction from the problem to the satisfiability problem.

Keywords: parameterized pattern matching, satisfiability, **NP**-complete

Different variants of the problem of the longest common subsequence are extensively used as distance measures for strings. In particular, the following problem was proposed in [1] (see also [2]). STR-IC-LCS:

Given two strings S_1 *and* S_2 *and a constraint pattern* P *of length* n, m, *and* r, *respectively, find a longest common subsequence of* S_1 *and* S_2 *including* P *as a substring.*

Another well-studied string comparison measure is that of parameterized matching (basic definitions and results can be found in [3]). It is natural to attempt to accommodate parameterized matching along with some other distance measures. In this paper we consider a parameterized variant of STR-IC-LCS.

THE PROBLEM OF THE LONGEST COMMON PARAMETERIZED SUBSEQUENCE WITH FIXED COMMON SUBSTRING (STR-IC-LCPS):

INSTANCE: *An alphabet* $\Sigma \cup \Pi$, *sequences* S_1 *and* S_2 *over* $\Sigma \cup \Pi$, *a string* P *over* Σ, *and positive integer* k.

QUESTION: *Is there a sequence* T, $|T| \geq k$, *such that* P *is a substring of* T *and* T *is a parameterized subsequence of* S_1 *and* S_2*?*

It is clear that there is some connection between longest common parameterized subsequences and longest common parameterized subsequences with fixed common substring. In particular, if T_1 is a longest common parameterized subsequence of S_1 and S_2 and T_2 is a parameterized subsequence of S_1 and S_2 with fixed common substring P, then $|T_1| \geq |T_2|$. However, T_1 and T_2 may significantly differ from each other.

Theorem 1. *For any* n *and* k, *there are sequences* S_1, S_2, P, T_1, *and* T_2 *such that*

(1) T_1 *is a longest common parameterized subsequence of* S_1 *and* S_2*;*

(2) T_2 *is a longest common parameterized subsequence of* S_1 *and* S_2 *with fixed common substring* P*;*

(3) $|T_2| \geq n$*;*

(4) $|T_1| \geq |T_2| + k$.

Proof. Let $\Sigma = \{a, b\}$, $\Pi = \emptyset$, $S_1 = a^s b^t$, $S_2 = b^t a^s$, $P = b^t$. We assume that $s > t + k$ and $t > n$. Let $T_1 = a^s$. Since $s > t + k$, it is clear that T_1 is a longest common parameterized subsequence of S_1 and S_2. Let $T_2 = b^t$. It is easy to see that T_2 is a parameterized subsequence of S_1 and S_2. Since $P = b^t$, it is clear that P is a substring of T_2. In view of $P = b^t$, it is easy to check that T_2 is a longest common parameterized subsequence of S_1 and S_2 with fixed common substring P. By definition of T_2, in view of $t > n$, it is clear that $|T_2| \geq n$. Since $s > t + k$, it is easy to see that $|T_1| \geq |T_2| + k$. □

Now we consider the complexity of STR-IC-LCPS.

Theorem 2. STR-IC-LCPS *is* **NP***-complete.*

Proof. It is clear that STR-IC-LCPS is in **NP**. In order to prove that STR-IC-LCPS is **NP**-hard, we shall reduce LCPS (see [4]) to STR-IC-LCPS.

LCPS:

INSTANCE: *An alphabet* $\Sigma \cup \Pi$, *sequences* S_1 *and* S_2 *over* $\Sigma \cup \Pi$, *and positive integer* k.

QUESTION: *Is there a sequence* T, $|T| \geq k$, *that is a parameterized subsequence of* S_1 *and* S_2*?*

Let $\Sigma \cup \Pi$ be an alphabet. Let S_1 and S_2 are sequences over $\Sigma \cup \Pi$.

We assume that c is a letter such that $c \notin \Sigma \cup \Pi$. Let $\Sigma' = \Sigma \cup \{c\}$. Let $P = c$ and $S_i' = cS_i$, $i \in \{1, 2\}$.

It is easy to check that T is a longest common parameterized subsequence of S_1 and S_2 if and only if cT is a longest common parameterized subsequence of S_1' and S_2' with fixed common substring P. Note that LCPS is **NP**-complete [4]. Therefore, STR-IC-LCPS is **NP**-complete. □

Encoding different hard problems as Boolean satisfiability and solving them with very efficient satisfiability algorithms has caused considerable interest (see e.g. [5] – [22]). We consider an explicit reduction from STR-IC-LCPS to the satisfiability problem.

Let $\Sigma = \{a_1, a_2, \ldots, a_{|\Sigma|}\}$, $\Pi = \{b_1, b_2, \ldots, b_{|\Pi|}\}$,

$$\varphi[1] = \wedge_{1\le i\le k} \vee_{1\le j\le|\Sigma\cup\Pi|}\, x[i,j],$$

$$\varphi[2] = \wedge_{1\le i\le k} \wedge_{1\le j[1]<j[2]\le|\Sigma\cup\Pi|}\, (\neg x[i,j[1]] \vee \neg x[i,j[2]]),$$

$$\varphi[3] = \wedge_{1\le i\le|P|} \vee_{1\le j\le|\Sigma|}\, u[i,j],$$

$$\varphi[4] = \wedge_{1\le i\le|P|} \wedge_{1\le j[1]<j[2]\le|\Sigma|}\, (\neg u[i,j[1]] \vee \neg u[i,j[2]]),$$

$$\varphi[5] = \wedge_{1\le i\le|P|} \wedge_{1\le j\le|\Sigma|, P[i]\ne a_j}\, \neg u[i,j],$$

$$\varphi[6] = \vee_{1\le i\le k-|P|+1} v[i],$$

$$\varphi[7] = \wedge_{1\le i\le k-|P|+1, 1\le j\le|P|, 1\le s\le|\Sigma|}((\neg v[i] \vee \neg u[j,s] \vee x[j+i-1,s])\wedge$$
$$(\neg v[i] \vee u[j,s] \vee \neg x[j+i-1,s])),$$

$$\varphi[8] = \wedge_{1\le i\le|S_2|} \vee_{1\le j\le|\Sigma\cup\Pi|}\, y[i,j],$$

$$\varphi[9] = \wedge_{1\le i\le|S_2|} \wedge_{1\le j[1]<j[2]\le|\Sigma\cup\Pi|}\, (\neg y[i,j[1]] \vee \neg y[i,j[2]]),$$

$$\varphi[10] = \wedge_{1\le i\le|S_2|, S_2[i]\in\Pi} \wedge_{1\le j\le|\Sigma|}\, \neg y[i,j],$$

$$\varphi[11] = \wedge_{\substack{1\le i[1]<i[2]\le|S_2|,\\ S_2[i[1]]=S_2[i[2]],\\ S_2[i[2]]\in\Pi,\\ 1\le j\le|\Sigma\cup\Pi|}}((\neg y[i[1],j] \vee y[i[2],j]) \wedge (y[i[1],j] \vee \neg y[i[2],j])),$$

$$\varphi[12] = \wedge_{\substack{1\le i[1]<i[2]\le|S_2|,\\ S_2[i[1]]\ne S_2[i[2]], S_2[i[1]]\in\Pi, S_2[i[2]]\in\Pi,\\ 1\le j\le|\Sigma\cup\Pi|}} (\neg y[i[1],j] \vee \neg y[i[2],j]),$$

$$\varphi[13] = \wedge_{1\le i\le|S_1|, 1\le j\le k, 1\le l\le|\Sigma\cup\Pi|, S_1[i]\ne a_l, S_1[i]\ne b_{l-|\Sigma|}}(\neg z[1,i,j] \vee \neg x[j,l]),$$

$$\varphi[14] = \wedge_{1\le i\le|S_2|, 1\le j\le k, 1\le l\le|\Sigma\cup\Pi|, S_2[i]\ne a_l, S_2[i]\in\Sigma}(\neg z[2,i,j] \vee \neg x[j,l]),$$

$$\varphi[15] = \wedge_{1\le i\le|S_2|, 1\le j\le k, 1\le l\le|\Sigma|, S_2[i]\in\Pi}(\neg z[2,i,j] \vee \neg x[j,l]),$$

$$\varphi[16] = \wedge_{1\le i\le|S_2|, 1\le j\le k, |\Sigma|+1\le l\le|\Sigma\cup\Pi|, S_2[i]\in\Pi}((\neg z[2,i,j] \vee \neg y[i,l] \vee x[j,l])\wedge$$
$$(\neg z[2,i,j] \vee y[i,l] \vee \neg x[j,l])),$$

$$\varphi[17] = \wedge_{1\le i\le 2, 1\le j\le|S_i|, 1\le l[1]<l[2]\le k}(\neg z[i,j,l[1]] \vee \neg z[i,j,l[2]]),$$

$$\varphi[18] = \wedge_{1 \le i \le 2, 1 \le l \le k} \vee_{1 \le j \le |S_i|} z[i, j, l],$$

$$\varphi[19] = \wedge_{\substack{1 \le i \le 2, 1 \le j[1] \le |S_i|, \\ 1 \le l[1] \le k, \\ j[2] > j[1], l[2] < l[1]}} (\neg z[i, j[1], l[1]] \vee \neg z[i, j[2], l[2]]),$$

$$\xi = \wedge_{i=1}^{19} \varphi[i].$$

It is easy to check that there is a sequence T, $|T| \ge k$, such that P is a substring of T and T is a parameterized subsequence of S_1 and S_2 if and only if ξ is satisfiable. It is clear that ξ is a CNF. So, ξ gives us an explicit reduction from STR-IC-LCPS to SAT. Now, using standard transformations (see e.g. [23]) we can obtain an explicit transformation ξ into ζ such that $\xi \Leftrightarrow \zeta$ and ζ is a 3-CNF. Clearly, ζ gives us an explicit reduction from STR-IC-LCPS to 3SAT.

We have designed generators of natural random instances for STR-IC-LCPS. We have consider our genetic algorithms OA[1] (see [24]), OA[2] (see [25]), OA[3] (see [26]), and OA[4] (see [27]) for SAT. We have used heterogeneous cluster. Each test was runned on a cluster of at least 100 nodes. Note that due to restrictions on computation time (20 hours) we used savepoints. Selected experimental results are given in Table 1.

time	average	max	best
OA[1]	2.7 h	31.72 h	14 min
OA[2]	2.32 h	26.4 h	18 min
OA[3]	1.74 h	29 h	26 min
OA[4]	1.96 h	14.4 h	21.2 min

Table 1: Experimental results for STR-IC-LCPS.

ACKNOWLEDGEMENTS. The work was partially supported by Analytical Departmental Program "Developing the scientific potential of high school" 8.1616.2011.

References

[1] Y.-C. Chen and K.-M. Chao, On the generaized constrained longest common subsequence problem, *Journal of Combinatorial Optimization*, 21 (2011), 383-392.

[2] E. Farhana, J. Ferdous, T. Moosa, and M.S. Rahman, Finite Automata Based Algorithms for the Generalized Constrained Longest Common Subsequence Problems, *Lecture Notes in Computer Science*, 6393 (2010), 243-249.

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

[4] O. Keller, T. Kopelowitz, and M. Lewenstein, On the Longest Common Parameterized Subsequence, *Theoretical Computer Science*, 410 (2009), 5347-5353.

[5] A. Gorbenko and V. Popov, On the Problem of Sensor Placement, *Advanced Studies in Theoretical Physics*, 6 (2012), 1117-1120.

[6] A. Gorbenko and V. Popov, On the Longest Common Subsequence Problem, *Applied Mathematical Sciences*, 6 (2012), 5781-5787.

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

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

[9] A. Gorbenko, V. Popov, and A. Sheka, Localization on Discrete Grid Graphs, *Lecture Notes in Electrical Engineering*, 107 (2012), 971-978.

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

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

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

[13] 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.

[14] 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.

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

[16] A. Gorbenko and V. Popov, Clustering Algorithm in Mobile Ad Hoc Networks, *Advanced Studies in Theoretical Physics*, 6 (2012), 1239-1242.

[17] A. Gorbenko and V. Popov, The Problem of Finding Two Edge-Disjoint Hamiltonian Cycles, *Applied Mathematical Sciences*, 6 (2012), 6563-6566.

[18] A. Gorbenko and V. Popov, Hamiltonian Alternating Cycles with Fixed Number of Color Appearances, *Applied Mathematical Sciences*, 6 (2012), 6729-6731.

[19] A. Gorbenko and V. Popov, Footstep Planning for Humanoid Robots, *Applied Mathematical Sciences*, 6 (2012), 6567-6571.

[20] A. Gorbenko and V. Popov, Multiple Occurrences Shortest Common Superstring Problem, *Applied Mathematical Sciences*, 6 (2012), 6573-6576.

[21] A. Gorbenko and V. Popov, The Far From Most String Problem, *Applied Mathematical Sciences*, 6 (2012), 6719-6724.

[22] A. Gorbenko and V. Popov, Multi-agent Path Planning, *Applied Mathematical Sciences*, 6 (2012), 6733-6737.

[23] A. Gorbenko and V. Popov, The c-Fragment Longest Arc-Preserving Common Subsequence Problem, *IAENG International Journal of Computer Science*, 39 (2012), 231-238.

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

[25] A. Gorbenko and V. Popov, Computational Experiments for the Problem of Selection of a Minimal Set of Visual Landmarks, *Applied Mathematical Sciences*, 6 (2012), 5775-5780.

[26] A. Gorbenko and V. Popov, Task-resource Scheduling Problem, *International Journal of Automation and Computing*, 9 (2012), 429-441.

[27] A. Gorbenko and V. Popov, SAT Solvers for the Problem of Sensor Placement, *Advanced Studies in Theoretical Physics*, 6 (2012), 1235-1238.

Received: November 1, 2012