Restricted Common Superstrings

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 restricted common superstring problem. This approach is based on an explicit reduction from the problem to the satisfiability problem.

Keywords: restricted common superstring, NP-complete, satisfiability

Computational complexity of different problems of finding regularities and efficient algorithms for these problems are thoroughly studied in theoretical computer science (see e.g. [1] – [6]). In particular, complexity of the restricted common superstring problem and some approximation algorithms for the problem was considered in [7, 8].

Let $\Sigma = \{a_1, \ldots, a_m\}$ be a finite alphabet. Let $S = \{S_1, \ldots, S_n\}$ be a set of strings over $\Sigma$. We assume that $S[i]$ is the $i$th letter in string $S$. Also, we assume that $S[i,j]$ is the substring of $S$ consisting of the $i$th letter through the $j$th letter. The length of a string $S$ is the number of letters in it. We assume that $|S|$ is the length of $S$. We use $\#oc(X,Y)$ to denote $|\{i \mid X = Y[i,j]\}|$. 
The decision version of the restricted common superstring problem can be formulated as following.

**The restricted common superstring problem (RCSstr):**

**Instance:** A set $S$ of strings over $\Sigma$, a string $T$, and a positive integer $k$.

**Question:** Is there a string $S$ such that $|S| = |T|$, $S[\pi(i)] = T[i]$, for all $1 \leq i \leq |T|$, and $|\{i \mid \#occ(S_i, S) \geq 1\}| \geq k$?

The problem RCSstr is $\text{NP}$-complete [7]. Note that encoding different hard problems as instances of SAT and solving them with efficient satisfiability algorithms has caused considerable interest (see e.g. [9] – [20]). In this paper, we consider an approach to solve the RCSstr problem. Our approach is based on an explicit reduction from the problem to the satisfiability problem.

Let

$$\varphi[1] = \land_{1 \leq i \leq |T|} \lor_{1 \leq j \leq |T|} w[i, j],$$
$$\varphi[2] = \land_{1 \leq i \leq |T|} \land_{1 \leq j_1 < j_2 \leq |T|} (\neg w[i, j_1] \lor \neg w[i, j_2]),$$
$$\varphi[3] = \land_{1 \leq i \leq |T|} \lor_{1 \leq j \leq m} x[i, j],$$
$$\varphi[4] = \land_{1 \leq i \leq |T|} \land_{1 \leq j_1 < j_2 \leq m} (\neg x[i, j_1] \lor \neg x[i, j_2]),$$
$$\varphi[5] = \land_{1 \leq i \leq |T|} \land_{1 \leq j \leq |T|} \land_{1 \leq p \leq m, T[j] \neq T[p]} (\neg w[i, j] \lor \neg x[i, p]),$$
$$\varphi[6] = \land_{1 \leq i \leq k} \land_{1 \leq j \leq n} y[i, j],$$
$$\varphi[7] = \land_{1 \leq i \leq k} \land_{1 \leq j_1 < j_2 \leq n} (\neg y[i, j_1] \lor \neg y[i, j_2]),$$
$$\varphi[8] = \land_{1 \leq i \leq k} (\neg y[i, j] \lor (\lor_{1 \leq p \leq |T| - |S|_j + 1} \land_{1 \leq j \leq n} y[j, p]),$$
$$\varphi[9] = \land_{1 \leq i \leq k} \land_{1 \leq p \leq |T| - |S|_j + 1} (\neg y[i, j] \lor \neg z[j, p[1]] \lor z[j, p[2]]),$$
$$\varphi[10] = \land_{1 \leq i \leq k} \land_{1 \leq r \leq m, S_j[q-p+1] \neq a_r} (\neg y[i, j] \lor \neg z[j, p] \lor \neg x[q, r]),$$

$$\xi = \land_{i=1}^{10} \varphi[i].$$

It is easy to check that there is a string $S$ such that $|S| = |T|$, $S[\pi(i)] = T[i]$, for all $1 \leq i \leq |T|$, and $|\{i \mid \#occ(S_i, S) \geq 1\}| \geq k$ if and only if $\xi$ is satisfiable. It is clear that $\xi$ is a CNF. So, $\xi$ gives us an explicit reduction from RCSstr to SAT. Now, using standard transformations (see e.g. [21]) we can obtain an explicit transformation $\xi$ into $\zeta$ such that $\xi \iff \zeta$ and $\zeta$ is a 3-CNF. Clearly, $\zeta$ gives us an explicit reduction from RCSstr to 3SAT.

We have designed generators of natural instances for RCSstr. We consider our genetic algorithms OA[1] (see [22]), OA[2] (see [23]), OA[3] (see [24]), and OA[4] (see [25]) for SAT. We used heterogeneous cluster. Each test was runned...
on a cluster of at least 100 nodes. Selected experimental results are given in Table 1.

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

### References


Received: February 12, 2013