The Minimum $k$-Cover 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

We consider the problem of determining the minimum cardinality collection of substrings, each of given length $k \geq 2$, that “cover” a given string $x$ of length $n$. We describe an approach to solve this problem. This approach is based on constructing an explicit reduction from the problem to the satisfiability problem.

Keywords: strings, $k$-covers, satisfiability

Different problems of finding regularities are thoroughly studied in theoretical computer science (see e.g. [1] – [6]). In particular, the minimum $k$-cover problem was introduced in [7].

Given a nonempty string $x$ of length $n$, a set $V = \{v_1, v_2, \ldots, v_p\}$ of $p$ substrings of $x$. We say that $V$ is a cover for $x$ if and only if every position of $x$ lies within an occurrence of some $v_i$, $1 \leq i \leq p$. In addition, if each string in $V$ has length $k$, then $V$ is a $k$-cover of $x$. If $p$ is the minimum integer for which such a set $V$ exists, then $V$ is said to be a minimum $k$-cover of $x$. 
The Minimum k-Cover Problem (MCP):

INSTANCE: An alphabet \( \Sigma \), a string \( X \) over \( \Sigma \), positive integers \( k \) and \( p \).

QUESTION: Whether there exists a \( k \)-cover of \( X \) of cardinality \( p \)?

The minimum \( k \)-cover problem is \( \text{NP} \)-complete (see [8]). Encoding problems as Boolean satisfiability and solving them with very efficient satisfiability algorithms has recently caused considerable interest (see e.g. [9] – [25]). In this paper, we consider an explicit reduction from MCP to the satisfiability problem. For simplicity, we use \( S[i] \) to denote the \( i \)th letter in sequence \( S \), and \( S[i, j] \) to denote the substring of \( S \) consisting of the \( i \)th letter through the \( j \)th letter. Let \( \Sigma = \{ a_1, a_2, \ldots, a_{|\Sigma|} \} \). Let

\[
\begin{align*}
\varphi[1, i, j] &= \lor_{1 \leq l \leq |\Sigma|} x[i, j, l], \\
\varphi[2, i, j] &= \land_{1 \leq l \leq |\Sigma|, 1 \leq |l| \leq |\Sigma|, l[1] \neq l[2]} (\neg x[i, j, l[1]] \lor \neg x[i, j, l[2]]), \\
\varphi[i, j] &= \varphi[1, i, j] \land \varphi[2, i, j], \\
\varphi &= \land_{1 \leq i \leq p, 1 \leq j \leq |\Sigma|} \varphi[i, j], \\
\psi[i] &= \lor_{1 \leq j \leq |X|} y[i, j], \\
\psi &= \land_{1 \leq i \leq p} \psi[i], \\
\rho[i] &= \lor_{1 \leq j \leq p, h_i \leq i \leq h_i = 1, \text{if } i \leq k, h_i = i - k + 1, \text{if } i > k} y[j, l], \\
\rho &= \land_{1 \leq i \leq |X|} \rho[i], \\
\tau[1, i] &= \land_{1 \leq j \leq |\Sigma|, X[i] = a_1, l \neq j} \neg z[i, j], \\
\tau[2] &= \land_{1 \leq i \leq |X|, X[i] = a_j} \neg z[i, j], \\
\tau &= \tau[2] \land \land_{1 \leq i \leq |X|} \tau[1, i], \\
\eta &= \land_{1 \leq i \leq p, 1 \leq j \leq |X| - k + 1} (\neg z[i, j] \lor \neg z[i, j] \lor \neg x[i, j, l[1]] \lor \neg x[i, j, l[2]]) \\
\xi &= \varphi \land \psi \land \rho \land \tau \land \eta.
\end{align*}
\]

**Theorem.** Given a fixed alphabet \( \Sigma \), a string \( X \) over \( \Sigma \), positive integers \( k \) and \( p \). There is a \( k \)-cover of \( X \) if and only if \( \xi \) is satisfiable.

**Proof.** Suppose that there is \( V = \{ v_1, v_2, \ldots, v_p \} \) that is a \( k \)-cover of \( X \) of cardinality \( p \). Let \( x[i, j, l] = 1 \) where \( 1 \leq i \leq p, 1 \leq j \leq k, v_i[j] = a_l; x[i, j, l] = 0 \) where \( 1 \leq i \leq p, 1 \leq j \leq k, v_i[j] = a_l; y[i, j] = 1 \) if and only if \( X[j, j + k - 1] = x[i] \) where \( 1 \leq i \leq p, 1 \leq j \leq |X| - k + 1; z[i, j] = 1 \) where \( 1 \leq i \leq |X|, 1 \leq j \leq |\Sigma|, X[i] = a_j; z[i, j] = 0 \) where \( 1 \leq i \leq |X|, 1 \leq j \leq |\Sigma|, X[i] \neq a_j \).

Since \( V \subseteq \Sigma^k \), for all \( i \) and \( j \) there is \( l \) such that \( x[i, j, l] = 1 \). Therefore, \( \varphi[1, i, j] = 1 \). In view of \( x[i, j, l] = 0 \) where \( 1 \leq i \leq p, 1 \leq j \leq k, v_i[j] \neq a_l \), it is clear that there is no more than one value of \( l \) such that \( x[i, j, l] = 1 \). Hence either \( x[i, j, l[1]] = 0 \) or \( x[i, j, l[2]] \) for all \( i, j, l[1] \neq l[2] \). Therefore, \( \varphi[2, i, j] = 1 \). So, \( \varphi = 1 \).
Note that $V$ is a set of substrings of $X$. Since $y[i, j] = 1$ if and only if $X[j, i + k - 1] = v_i$, it is easy to see that $\psi[i] = 1$. By definition, $\psi[2, i] = 1$. So, $\psi = 1$.

Since $V$ is a $k$-cover of $X$, $X[r, r + k - 1] = v_j$ for some $r$ and $j$ such that $1 \leq j \leq p$, $r \leq i \leq r + k - 1$. Therefore, $\rho[i] = 1$. So, $\rho = 1$. Since $z[i, j] = 1$ where $1 \leq i \leq |X|$, $1 \leq j \leq |\Sigma|$, $X[i] = a_j$; $z[i, j] = 0$ where $1 \leq i \leq |X|$, $1 \leq j \leq |\Sigma|$, $X[i] \neq a_j$, it is easy to check that $\tau = 1$. Since $V$ is a $k$-cover of $X$, it is clear that $\eta = 1$. Therefore, $\xi = 1$.

Suppose now that $\xi = 1$. Hence $\xi = \varphi = \psi = \rho = \tau = \eta = 1$. Since $\varphi = 1$, by definition, $\varphi[1, i, j] = 1$, $\varphi[2, i, j] = 1$. It is easy to check that for all $i$ and $j$ there is only one value of $l$ such that $x[i, j, l] = 1$. Let $v_i[j] = a_i$. Since $\eta = 1$ and $\tau = 1$, it is clear that if $y[i, j] = 1$, then $X[j, i + k - 1] = v_i$. In view of $\rho = 1$, we obtain that $V$ is a $k$-cover of $X$. □

In view of the theorem, we obtain an explicit reduction from MCP to PSAT.

Note that $\alpha \rightarrow \beta \iff \neg\alpha \lor \beta$, $\alpha = \beta \iff (\neg\alpha \lor \beta) \land (\alpha \lor \neg\beta)$. Therefore, $\eta \iff \eta'$ where

$$\eta' = \land_{1 \leq i \leq p, 1 \leq j \leq |X|} \land_{-k+1,1 \leq t \leq k-1,1 \leq l \leq |\Sigma|}(\neg y[i, j] \lor \neg z[j + t, l] \lor x[i, 1 + t, l]) \land \neg y[i, j] \lor z[j + t, l] \lor \neg x[i, 1 + t, l]).$$

Let $\xi' = \varphi \land \psi \land \rho \land \tau \land \eta'$. It is clear that $\xi \iff \xi'$. Since $\xi'$ is a CNF, we obtain an explicit reduction from MCP to SAT.

Using standard transformations (see e.g. [26]) we can obtain an explicit transformation $\xi'$ into $\xi''$ such that $\xi' \iff \xi''$ and $\xi''$ is a 3-CNF. It is easy to see that $\xi''$ gives us an explicit reduction from MCP to 3SAT.

There is a well known site on which posted solvers for SAT [27]. They are divided into two main classes: stochastic local search algorithms and algorithms improved exhaustive search. All solvers allow the conventional format for recording DIMACS boolean function in conjunctive normal form and solve the corresponding problem [28]. 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 create a generator of natural instances for LCS. Also we use test problems from [27]. We use algorithms from [27]. Also we design our own genetic algorithm for SAT which based on algorithms from [27].

We use 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) [29].

Each test was run on a cluster of at least 100 nodes. The maximum solution time was 6 hours. The average time to find a solution was 11.4 minutes. The best time was 7 seconds.
ACKNOWLEDGEMENTS. The work was partially supported by Analytical Departmental Program “Developing the scientific potential of high school” 8.1616.2011.

References


The minimum \( k \)-cover problem


Received: February 12, 2013