Information Processing Letters 26 (1987/88) 269–270
North-Holland

11 January 1988

NOTE

SHORT PROPOSITIONAL FORMULAS REPRESENT NONDETERMINISTIC COMPUTATIONS

Stephen A. COOK

Department of Computer Science, University of Toronto, Toronto, Ontario, Canada M5S 1A4

Communicated by T. Lengauer
Received 10 February 1987
Revised 20 May 1987

Keywords: NP-complete, satisfiability, nondeterministic Turing machine

In this note, n refers to the length of the input x of a machine and the *length* of a propositional formula refers to the number of occurrences of atoms in it.

Theorem. *Let* M *be a* T(n) *time-bounded nondeterministic multitape Turing machine. For each input* x *there is a conjunctive normal form propositional formula* F(x) *of length* O(T(n) log T(n)) *such that* F(x) *is satisfiable if and only if* M *accepts* x *within* T(n) *steps.*

A weaker form of the above result first appeared in [1], where it was used to show that CNF satisfiability is NP-hard. The theorem in [1] referred only to single-tape machines, and the length bound on the formula was something like $T(n)^3$. The length bound was improved to T(n) log T(n) by Robson [4]. Recently, Stearns and Hunt [6] have given a different construction which yields a slightly longer formula, namely one of length $O(T(n) \log^2 T(n))$, but which applies to multitape Turing machines. All of the above constructions are fairly elaborate.

Our theorem above can be proved by the methods of Robson [4]. The proof is also implicit in Schnorr's proof [5] that every language accepted by a nondeterministic multitape Turing machine in quasilinear time (i.e., within $n \log^k n$ steps for some $k \geq 1$) is quasilinear time reducible to the satisfiability problem. Our purpose is to state the result explicitly in its strong form, and point out that the hard part of the proof is already contained in a standard complexity theory result.

The theorem is interesting because it gives information about the relative complexity of different NP problems. For example, Schnorr [5] uses it to show that SAT is quasilinear complete, and hence so are a number of other standard NP-complete problems. Stearns and Hunt [6] use it to show that, assuming SAT has asymptotic complexity 2^n, many standard NP-complete problems have about the same complexity as SAT, but others, such as the CLIQUE problem, are easier. Dewdney [2] is interested because of its potential application to his generic reduction computer, which implements nondeterministic algorithms.

To prove the theorem, note that we can convert M to a deterministic machine M_1 by providing a second binary input tape I_2 which specifies the nondeterministic choices M is to make during its computation. Thus, M accepts an input x iff there is some bit string for I_2 of length T(n) such that M_1 accepts x. Thus, by the construction in [3], for each input length n there is a Boolean circuit C_n of size O(T(n) logT(n)) which accepts the same input pairs (x, I_2) as M_1 does. Now for each input x of length n we construct the CNF formula F(x) by introducing, for each gate or input node g of the circuit C_n, an atom P_g representing the value of g.

For each (non-input) gate g, clauses with two or three literals are introduced which express P_g in terms of the atoms representing the inputs of g. If g represents one of the input bits of x, then P_g is replaced by either 0 or 1, according to the value of that bit in x, and the usual simplications are made to keep the formula in conjunctive normal form. Finally, if g is the output gate of the circuit, then the unit clause P_g is introduced. The resulting formula has a fixed number of clauses of length at most three for each gate in C_n, and hence the total length of F(x) is O(T(n) log T(n)).

Any truth assignment satisfying F(x) gives the correct values to all gates in the circuit C_n for input x and some input I_2, and hence C_n has output 1 and M accepts x. Conversely, if M accepts x, then there is a satisfying assignment for F(x) corresponding to an appropriate accepting computation of C_n.

An interesting open question is whether the length of the formula F(x) can be shortened, say to O(T(n)).

References

[1] S.A. Cook, The complexity of theorem-proving procedures, Proc. 3rd Ann. ACM Symp. on Theory of Computing (1971) 151–158.

[2] A.K. Dewdney, Generic Reduction Computers, Rept. No. 141, Dept. of Computer Science, Univ. of Western Ontario, October 1985.

[3] N. Pippenger and M.J. Fischer, Relations among complexity measures, J. ACM 26 (2) (1979) 361–381.

[4] J.M. Robson, A new proof of the NP completeness of satisfiability, Proc. 2nd Australian Computer Science Conf. (1979) 62–69.

[5] C.P. Schnorr, Satisfiability is quasilinear complete in NQL, J. ACM 25 (1) (1978) 136–145.

[6] R.E. Stearns and H.B. Hunt, III, On the complexity of the satisfiability problem and the structure of NP, Tech. Rept. 86-21, Dept. of Computer Science, State Univ. of New York at Albany, October 1986.