FINDING A HOMOMORPHISM BETWEEN TWO WORDS IS NP-COMPLETE

by

Andrzej Ehrenfeucht
Department of Computer Science
University of Colorado at Boulder
Boulder, Colorado

Grzegorz Rozenberg
Department of Mathematics
University of Antwerp, U.I.A.
Wilrijk, Belgium

#CU-CS-134-78 September, 1978

ALL correspondence to second author.

- WORKING PAPER -

Abstract. We demonstrate that to find a homomorphism between two words x and y where letters of x can be chosen from an infinite alphabet and y is a word over a two letter alphabet is an NP-complete problem.

Looking for NP-complete problems today forms an active research area within complexity theory (see, e.g., [1]). Showing that a problem is NP-complete often contributes to our understanding of the difficulty of the problem and of the nature of NP-complete problems in general.

In this note we demonstrate that a very common problem from formal language theory is NP-complete: finding a nonerasing homomorphism between two words x and y (providing we do not restrict a priori the size of the alphabet of the word x; y can be chosen over a two-letter alphabet.) The problem remains NP-complete even when arbitrary homomorphisms are admitted. It may very well be one of the simplest NP-complete problems known. On the other hand one can easily see that if the size of the alphabet is limited a' priori then the problem is in P.

Formally it is defined as follows. (In the sequel, given a word x, *alph* x denotes the set of letters appearing in x and $\#_a x$ denotes the number of occurrences of the letter a in x; $\mathrm{HOM}(\Sigma_1,\Sigma_2)$ denotes the class of nonerasing homomorphisms from Σ_1^* into Σ_2^*).

Let Σ be an infinite alphabet, Δ its subset containing two elements, $\Delta = \{b,¢\}$ say. Let

$$\mathrm{MATCH}(\Sigma,\Delta) = \{(x,y) : \mathit{alph}\, x \subset \Sigma,\ \mathit{alph}\, y \subseteq \Delta \text{ and there exists a homomorphism } h \text{ in } \mathrm{HOM}(\mathit{alph}\, x, \mathit{alph}\, y) \text{ such that } h(x) = y\}.$$

Theorem. Membership in $\mathrm{MATCH}(\Sigma,\Delta)$ is NP-complete.

Proof.

(i) Obviously membership in $\mathrm{MATCH}(\Sigma,\Delta)$ is in NP.

(ii) Since 3-satisfiability is NP-complete (see, e.g., [1]) it suffices to show that for every Boolean expression Ψ in 3-conjunctive normal form (3-CNF) there exist a word x_Ψ in Σ^+ and y_Ψ in Δ^+ such that

(*). . . $(x_\Psi,y_\Psi) \in \mathrm{MATCH}(\Sigma,\Delta)$ if and only if Ψ is satisfiable.

To this aim we proceed as follows.

Let $V = \{P_1,\ldots,P_\ell\}$ be a set of Boolean variables and let

$\Psi = (P'_{i_1} \vee P'_{j_1} \vee P'_{k_1}) \wedge \ldots \wedge (P'_{i_n} \vee P'_{j_n} \vee P'_{k_n})$,

with $i_q, j_q, k_q \in \{1,\ldots,\ell\}$ be a Boolean expression over V in 3-CNF where, for $q \in \{1,\ldots,n\}$, each P'_{i_q} (P'_{j_q}, P'_{k_q} respectively) is either a variable P_{i_q} (P_{j_q}, P_{k_q} respectively) or its negation $\overline{P}_{i_q}$ ($\overline{P}_{j_q}, \overline{P}_{k_q}$ respectively).

Our construction of words x_Ψ, y_Ψ takes two steps.

STEP 1.

We will construct a finite set W of pairs of words (α,β) with $\alpha \in V_\Psi^+, V_\Psi \subset \Sigma$, and $\beta \in \Delta^+$ such that

(**) . . . $\begin{cases} \Psi \text{ is satisfiable if and only if there exists a homomorphism} \\ h \text{ in } HOM(V_\Psi,\{b\}) \text{ such that, for every } (\alpha,\beta) \text{ in } W,\ h(\alpha) = \beta. \end{cases}$

Let $\overline{V} = \{\overline{P}_q : 1 \le q \le \ell\}$. Clearly we can assume that $V \cup \overline{V} \subseteq \Sigma$.

Let $T_1,\ldots,T_n$, $U_1,\ldots,U_n$ be new elements of Σ different from each other.

Let $V_\Psi = V \cup \overline{V} \cup \{T_q : 1 \le q \le n\} \cup \{U_q : 1 \le q \le n\}$ and let

$W_1 = \{(P_q\overline{P}_q, b^3) : 1 \le q \le \ell\}$,

$W_2 = \{(T_q P'_{i_q} P'_{j_q} P'_{k_q}, b^6) : 1 \le q \le n\}$

$W_3 = \{(T_q U_q, b^4) : 1 \le q \le n\}$.

Let $W = W_1 \cup W_2 \cup W_3$.

We prove (**) as follows.

(1) Assume that there exists a homomorphism h in $HOM(V_\Psi,\{b\})$ such that, for every (α,β) in W, $h(\alpha) = \beta$.

Then let f be the valuation of V such that for every $q \in \{1,\ldots,\ell\}$, $f(P_q) =$ *false* if and only if $h(P_q) = b^2$. Since $W_1 \subseteq W$, f is well defined (and it follows that $f(P_q) =$ *true* if and only if $h(P_q) = b$).

Since $W_3 \subseteq W$, $1 \le |h(T_q)| \le 3$ for $1 \le q \le n$ and, because $W_2 \subseteq W$, this implies that $3 \le |h(P'_{i_q} P'_{j_q} P'_{k_q})| \le 5$ for $1 \le q \le n$. Thus for every $q \in \{1,\ldots,n\}$ either $|h(P'_{i_q})| = 1$ or $|h(P'_{j_q})| = 1$ or $|h(P'_{k_q})| = 1$ which implies that for every $q \in \{1,\ldots,n\}$, $f((P'_{i_q} \vee P'_{j_q} \vee P'_{k_q})) =$ *true* and so Ψ is satisfied by f.

(2) Let Ψ be satisfiable and let f be a valuation of V which satisfies Ψ. Let h be the homomorphism on V_Ψ^* defined by: for $1 \le q \le \ell$,

if $f(P_q) =$ *true* then $h(P_q) = b$ and $h(\overline{P}_q) = b^2$,

if $f(P_q) =$ *false* then $h(P_q) = b^2$ and $h(\overline{P}_q) = b$,

for $1 \le q \le n$,

if all three of $f(P'_{i_q})$, $f(P'_{j_q})$, $f(P'_{k_q})$ are equal *true*

then $h(T_q) = b^3$ and $h(U_q) = b$,

if only two of $f(P'_{i_q})$, $f(P'_{j_q})$, $f(P'_{k_q})$ are equal *true*

then $h(T_q) = b^2$ and $h(U_q) = b^2$,

if only one of $f(P'_{i_q})$, $f(P'_{j_q})$, $f(P'_{k_q})$ is equal *true*

then $h(T_q) = b$ and $h(U_q) = b^3$.

It follows directly from the definition of h that indeed for every (α,β) in W, $h(\alpha) = \beta$.

Thus (**) holds.

STEP 2.

Now given W from STEP 1 we construct x_Ψ, y_Ψ satisfying (*) as follows Let $W = \{(\alpha_1,\beta_1),\ldots,(\alpha_m,\beta_m)\}$ for some $m \ge 3$ and let $x_\Psi = ¢\ \alpha_1\ ¢\ldots¢\ \alpha_m$ and $y_\Psi = ¢\ \beta_1\ ¢\ldots¢\ \beta_m$.

Note that if h is a homomorphism from *alph* $x_\Psi = V_\Psi \cup \{¢\}$ into *alph* $y_\Psi = B$ such that $h(x_\Psi) = y_\Psi$ then $h(¢) = ¢$ (because both x_Ψ and y_Ψ

start with ¢ and ¢ $\notin$ *alph* β_1). Since $\#_{¢}x_\Psi = \#_{¢}y_\Psi$ this implies that

(***) . . . there exists a homomorphism h from *alph* x_Ψ into *alph* y_Ψ such that $h(x_\Psi) = y_\Psi$ if and only if there exists a homomorphism g from V_Ψ into {b} such that $g(\alpha_q) = \beta_q$ for $1 \le q \le m$.

But (***) and (**) imply (*) and so the theorem holds.

Remark. If we change the definition of MATCH(Σ,Δ) to the definition of $\text{MATCH}_\Lambda(\Sigma,\Delta)$ by allowing arbitrary rather than only nonerasing homomorphisms then the theorem still remains true. That is we get the result: "The membership in $\text{MATCH}_\Lambda(\Sigma,\Delta)$ is NP-complete." The main idea of the proof is the same and the only changes to be made are the following ones:

(1) For every (α,β) in W_1 set $\beta = b$ (rather than $\beta = b^3$).

(2) For every (α,β) in W_2 set $\beta = b^3$ (rather than $\beta = b^6$)

(3) For every (α,β) in W_3 set $\beta = b^2$ (rather than $\beta = b^4$)

(4) Given a homomorphism h "satisfying" W set the valuation f of V by: $f(P_q) =$ *true* if and only if $h(P_q) = b$.

(5) Given a valuation f of V satisfying Ψ set the homomorphism h by

if $f(P_q) =$ *true* then $h(P_q) = b$ and $h(\overline{P}\ ) = \Lambda$,

if $f(P_q) =$ *false* then $h(P_q) = \Lambda$ and $h(\overline{P}_q) = b$,

if all three of $f(P'_{i_q})$, $f(P'_{j_q})$, $f(P'_{k_q})$ are equal *true*

then $h(T_q) = \Lambda$ and $h(U_q) = b^2$,

if only two of $f(P'_{i_q})$, $f(P'_{j_q})$, $f(P'_{k_q})$ are equal *true*

then $h(T_q) = b$ and $h(U_q) = b$,

if only two of $f(P'_{i_q})$, $f(P'_{j_q})$, $f(P'_{k_q})$ are equal *true*

then $h(T_q) = b^2$ and $h(U_q) = \Lambda$.

(6) In *STEP 2* define

$x_\Psi = ¢\ ¢\ \alpha_1\ ¢ \ldots ¢\ \alpha_m\ ¢\ \alpha_1\ ¢ \ldots ¢\ \alpha_m$ and

$y_\Psi = ¢\ ¢\ \beta_1\ ¢ \ldots ¢\ \beta_m\ ¢\ \beta_1\ ¢ \ldots ¢\ \beta_m$.

REFERENCES

[1] A. Aho, J. Hopcroft and J. Ullman, The Design and Analysis of Computer Algorithms, Addison-Wesley, Reading, Mass., 1974.

Acknowledgments. The second author is very much indebted to IBM Belgium and NFWO foundation for supporting his stay at the University of Colorado at Boulder which made the work on this and other papers possible.