18 Feb 1996

R-By-C Crozzle: An NP-Hard Problem

Michelle Gower

Ralph W. Wilkerson
Missouri University of Science and Technology, ralphw@mst.edu

Recommended Citation

M. Gower and R. W. Wilkerson, "R-By-C Crozzle: An NP-Hard Problem," *Proceedings of the ACM Symposium on Applied Computing*, pp. 73 - 76, Association for Computing Machinery, Feb 1996.
The definitive version is available at https://doi.org/10.1145/331119.331148

R-BY-C CROZZLE: AN NP-HARD PROBLEM

Michelle Gower, Ralph Wilkerson
Department of Computer Science,
University of Missouri-Rolla

KEYWORDS: Crozzle, NP-hard, complexity

ABSTRACT

In an Australian magazine, a monetary prize is awarded to the person with the best answer to a word puzzle called a Crozzle. The placement of words into a ten by fifteen grid obtaining the highest score is the best answer. Various search techniques have been employed to solve this problem, yet no one has shown whether there is a polynomial-time algorithm to find the best Crozzle. This paper creates a similar word puzzle, called R-by-C Crozzle, by lifting the constraint on the grid size. R-by-C Crozzle is not in NP, but there exists a polynomial reduction to it from the exact 3-set cover problem. Thus the R-by-C Crozzle is NP-hard. This paper also explores any implications that the complexity of the R-by-C Crozzle might have on the complexity of the original problem.

INTRODUCTION

The Crozzle is a word puzzle that appears in The Australian Women's Weekly magazine. The object of the game is to place words from a given list into a ten by fifteen empty grid such that the score of the arrangement is maximal. Only words from the current word list can be used, and they can appear at most once in the grid. The resulting structure of words must be one interlocking grid. This means that no word can be free-standing. Also, a blank must separate words in a single column or row. The scoring rules give each word appearing in the answer ten points, and each letter in the intersection of two words scores additional points according to Table I.

Table I Scores for intersection letters

a, b, c, d, e, f	2
g, h, i, j, k, l	4
m, n, o, p, q, r	8
s, t, u, v, w, x	16
y	32
z	64

Different search algorithms have been utilized and modified to find the best arrangement [2,3]; however, no one has established the complexity of the Crozzle problem. Problems are classified by their computational complexity into sets which include P, NP, NP-complete and NP-hard. The following are brief descriptions of these well-known classes. For a rigorous mathematical description of complexity theory, the reader is referred to Garey and Johnson's work [1] in which they use Turing machines as the model of computation.

P is the class of problems that have polynomial-time algorithms to find the solutions, whereas problems in NP must have polynomial-time algorithms to check the validity of possible solutions. It is easily shown that P is subset of NP, which means that a problem that has been shown not to be in NP will never have a polynomial time solution.

For one problem π' to be "polynomially reducible" to another problem π (symbolized by $\pi' \propto \pi$), there must be an answer-preserving polynomial-time algorithm that takes an instance of π' and converts it into an instance of π.

The set of NP-complete problems is the set of problems π in NP such that $\pi' \propto \pi$ for all problems π' in NP. Another class of problems is the set of all π for which there exists a polynomial reduction from some NP-complete problem to π. Then each π has at least the same order of steps as the NP-complete problems. Accordingly these problems are called NP-hard.

R-BY-C CROZZLE

It is easily shown that the Crozzle is in NP; however, the fixed size of the grid makes it difficult to find a reduction for an NP-completeness proof. To overcome this obstacle, a new problem, called an R-by-C Crozzle, is created by allowing the size of the grid to be additional inputs.

R-by-C Crozzle Is Not In NP

Unlike the original problem, the R-by-C Crozzle is not in NP. To make sure that the possible solution satisfies the constraints, the minimum amount of work required is an examination of each square (i.e. on the order of RxC). The number of steps is dependent upon the values of R and C rather than the size of the inputs. Since there is no relationship between RxC and the number (n) of words in the list, there cannot be a polynomial-time algorithm to check possible solutions for all values of R, C, and n. Therefore R-by-C Crozzle is not in NP.

Reduction From an Exact Cover By 3-Sets

There exists a polynomial reduction from the exact cover by 3-sets problem to the RxC Crozzle problem. Given a set T, a 3-set contains exactly 3 elements of T. The input to the exact cover by 3-sets problem is a set S, a family F containing n 3-sets of S, and an integer $m \leq n$. An exact 3-set cover of S is a subfamily with m 3-sets of S such that every element of S appears exactly once in the subfamily.

By definition, the reduction must take an instance of exact 3-set cover and convert it to an instance of the R-by-C Crozzle. The basic idea behind the reduction is that special words of length 3m corresponding to each 3-set in the cover will appear horizontally in the solution to the R-by-C Crozzle problem. A summary of the reduction steps appears in Table II.

Table II Reduction Algorithm

1. Let R = m, C = 3m.
2. For all n 3-sets in family, create w:
   Say 3-set = $\{s_h, s_i, s_j\}$ then
   let w_h = "s", w_i = "t", and w_j = "v".
   let $w_{x \neq h,i,j}$ = "a" where $1 \leq x \leq 3m$.
3. For each z, $1 \leq z \leq m$, create w:
   let w_z = "s",
   let $w_{x \neq z}$ = "a" where $1 \leq x \leq m$,
   Repeat the process for
   w_z = "t", and w_z = "v".
4. Let $k = 6m^2 + 82m$.

Let each row in the grid correspond to a 3-set in the covering subfamily and each column represent one element of S. Thus, the grid will need to be m by 3m. The reduction needs to construct the word list for the R-by-C Crozzle problem. Create a word w of length 3m from each of the n given 3-sets. Let each position in w (denoted as w_x for $1 \leq x \leq 3m$) correspond to a certain element of S(i.e. w_x corresponds to s_x, for $1 \leq x \leq 3m$). If s_x is not an element of the 3-set, then let w_x equal the letter "a". Otherwise a non-"a" letter in w_x will denote that element s_x is in the 3-set. Each 3-set contains three elements, say s_h, s_i, and s_j, and without loss of generality let $h < i < j$. Let w_h = "s", w_i = "t", and w_j = "v". The reasons behind choosing the letters "a", "s", "t", and "v" are explained below in the discussion of the score construction. This process creates n unique words of length 3m.

The important part of the exact cover problem is that each element of S appears exactly once in the cover. Since non-"a" letters denote that an element is in a set, only one such letter appearing in a column means that an element appears once in the cover. To this end, words of length m are created such that for each z, $1 \leq z \leq m$, w_z = "s" and $w_{x \neq z}$ = "a", $1 \leq x \leq m$. Repeat this process for w_z = "t", and w_z = "v". This process creates 3m unique words of length m making the total numbers of words equal to n + 3m.

The reduction utilizes the score to force the resulting cover to contain exactly m 3-sets and each element of S to appear exactly once in the cover. If all m rows contain words of length 3m, then the cover will exactly contain m 3-sets. Each element in S will appear exactly once in the cover if and only if every column contains an m-length word with only one non-"a" letter. Thus the solution must contain m horizontal words and 3m vertical words. There will be $3m^2$ letters of which only 3m will be non-"a"s. The non-"a" letters are more significant, so they must score more in the solution. This is the reason for the choice of letters "a", "s", "t", and "v". The score of the Crozzle is 10 * (number of words) + (score of each intersecting letter). Therefore the score is equal to $10(m + 3m) + 16(3m) + 2(3m^2 - 3m) = 6m^2 + 82m$. So for the reduction define the minimum score (k) as $6m^2 + 82m$.

Answer Preserving

The reduction must be answer preserving, which means that there exists a subfamily of m 3-sets that exactly covers S if and only if there is an interlocking structure of words that scores $6m^2 + 82m$ or more.

<u>If</u>

Suppose there is a valid interlocking structure of words that scores $6m^2 + 82m$ or more. Then the grid must be completely full. Suppose this is not the case. Consider the maximum score of this solution. Let κ, ρ_m, ρ_{3m}, ι, τ be defined as in Table III.

Table III Variable Definitions

Let	κ	= Number of vertical words
	ρ_{3m}	= Number of horizontal words of length 3m
	ρ_m	= Number of horizontal words of length m
	ι	= Number of interlocking letters
	τ	= Number of interlocking letters s, t, or v

The Crozzle can have at most m vertical words in consecutive columns. Otherwise the vertical words would result in horizontal words of length other than m which are not in the word list. Since the number of columns is 3m, the Crozzle can maximally contain two such sets of m consecutive vertical words. In the remaining part of the Crozzle the vertical words must occur in alternating columns. Thus the maximum number of vertical words (κ) is $2m+\left\lfloor \frac{m-1}{2} \right\rfloor$.

The number (ρ_{3m}) of horizontal words of length 3m is limited to alternating rows and thus cannot exceed $\left\lceil \frac{m}{2} \right\rceil$. The remaining rows can contain two m-length words in the spaces occupied by the consecutive columns. So the number of horizontal words of length m (ρ_m) has an upper bound of $2(m-\rho_{3m})$.

The two sets of m consecutive vertical words contribute $2m^2$ letters in word intersections to the Crozzle. The alternating vertical words contribute such a letter only when they intersect with a horizontal word of length 3m. Thus the total number of intersections (ι) is constrained by $2m^2+\left\lfloor \frac{m-1}{2} \right\rfloor \rho_{3m}$. Only one intersection in a column can be an "s", "t", or "v" due to the construction of the m-length words. This limits the number (τ) of word intersections containing a non-"a" letter to the number of vertical words. However, the number of such letters in each row depends upon the horizontal words. Each m-length horizontal words contributes one non-"a" letter, whereas each 3m-length word has 3 non-"a" letters. The total number of non-"a" letters due to the words that lie horizontally equals $\rho_m + 3\rho_{3m} = 2m + \rho_{3m}$. Therefore the upper bound on τ is $2m+\min\left(\left\lfloor \frac{m-1}{2} \right\rfloor, \rho_{3m}\right)$.

Since this Crozzle contains only the letters "a", "s", "t", and "v", the score equals 10(number of words) + 2(number of intersections containing "a") + 16(number of intersections containing "s", "t", or "v"). Thus

$$score = 10(\kappa+\rho_{3m}+\rho_m)+2(\iota-\tau)+16\tau \qquad (1)$$

The maximum score of this Crozzle is obtained by substituting the upper bound of each variable. Since the 3m-length words contribute more intersections in the Crozzle, the score will be the largest when the number of 3m-length words rather than the m-length words is maximized. The result of making substitutions of the upper bounds into equation (1) and performing some simplification is:

$$score = 4m^2+68m+24\left\lfloor \frac{m-1}{2} \right\rfloor - 10\left\lceil \frac{m}{2} \right\rceil + 2\left\lfloor \frac{m-1}{2} \right\rfloor \left\lceil \frac{m}{2} \right\rceil \qquad (2)$$

But $\lfloor x \rfloor \leq x$ and $\lceil x \rceil < x + 1$ so equation (2) can be rewritten as $score \leq 4.5m^2+75m$. This contradicts the original assumption that the Crozzle scored $6m^2 + 82m$ or more, and thus the grid is completely full.

The subset that corresponds to each horizontal word will be placed in the subfamily. Because of the way these words were constructed in step two, there will be one "s", one "t", and one "v" in each word, say w(h), w(i), and w(j). Then the corresponding 3-set is $\{s_h, s_i, s_j\}$ (which exists also due to step two). The grid contains m rows which leads to exactly m subsets. Each column contains only one non-"a" letter due to the way the m-length words were created in step three. Since each position in the 3m-length words corresponds to an element of S, this means that each s_i appears exactly once in the subfamily. Therefore there exists a subfamily of m subsets that exactly covers S.

<u>Only If</u>

Given an exact 3-set cover, the following section proves that there exists a placement of the words that scores $6m^2 + 82m$ or more. It follows directly from the construction of the reduction. Suppose there exists a subfamily of m 3-sets that exactly covers S with each 3-set containing three elements, say s_h, s_i, s_j, where $h < i < j$. Place as a horizontal word w where w_h = "s", w_i = "t", w_j = "v", and $w_{x \neq h,i,j}$ = "a" for $1 \leq x \leq 3m$. These words are in the word list due to step two in the reduction. The m 3-sets result in m unique words being placed horizontally in the grid. For the Crozzle to be valid, the vertical words also must be unique and exist in the word list. The 3m-length words were constructed such that each position in the word corresponded to an element in S. Because the subfamily exactly covers S, each element in S occurs exactly once in the subfamily and thus in only one 3-set. For element s_x, consider the horizontal word that contains s_x. It will contain an "s", "t", or "v" in position x. All other horizontal words must contain an "a" in position x since the corresponding 3-sets do not contain s_x. This means that there is exactly one non-"a" letter in each position and thus each vertical word. These m-length words were placed in the word list by reduction step number three. Assume the vertical words are not unique. Then at least two words are equal, say w and x. So $w = x = w_1w_2...w_m$ and without loss of generality say $w_h =$ "s". Then some horizontal word, say z, equals z_1..."s"..."s"...z_{3m} which is a contradiction. The score of this Crozzle equals $10(m + 3m) + 16(3m) + 2(3m^2 - 3m)$ which reduces to $6m^2 + 82m$. Therefore there exists an interlocking structure of words that scores $6m^2 + 82m$ or more.

EFFECTS OF THIS COMPLEXITY

All instances of the Crozzle problem are contained in the set of the instances of the R-by-C Crozzle problem. Now that the R-by-C Crozzle problem has been proven to be NP-hard, one might ask what can be said about the complexity of the original Crozzle problem. Unfortunately, the complexity of the R-by-C Crozzle only provides an upper bound for the complexity of the Crozzle problem. An example of a problem that does not inherit the complexity of a superset occurs within the satisfiability problem (SAT). The subset of SAT with all instances of clauses containing exactly two elements is called 2-SAT. Cook proved that SAT is NP-complete; however, there is a polynomial time solution for 2-SAT [1]. Therefore, the NP-hardness proof of R-by-C Crozzle problem only implies that the complexity of the Crozzle problem can be no greater than NP-hard.

CONCLUSION

Complexity theory classifies algorithms according to the upper bound on the number of steps for a given input size. Basically, there are four well-known groups of algorithms: P, NP, NP-complete, and NP-hard. The Crozzle problem is a word puzzle that involves placing words in a grid such that the resulting score is more than some specified value. It is in NP, but a reduction to the Crozzle is difficult to find due to the fixed size of the grid. The R-by-C Crozzle problem was created by modifying the Crozzle problem so that the size of the grid is given as input. R-by-C Crozzle is not in NP which means that there will never be a polynomial-time algorithm that solves a R-by-C Crozzle. The exact 3-set cover problem can be polynomially reduced to the R-by-C Crozzle problem. This classifies the R-by-C Crozzle as NP-hard. Unfortunately, this result only places an upper bound on the complexity of the actual Crozzle problem.

ADDRESSES

The authors can be contacted at:
Michelle Gower
Department of Computer Science
University of Missouri-Rolla
Rolla, MO 65401
email: mgower@cs.umr.edu

REFERENCES

[1] Garey, M.R., Johnson, D.S., *Computers and Intractability: A Guide to the Theory of NP-completeness*, New York: W.H. Freeman and Company, 1979.

[2] Harris G.H., Forster J.J.H., "Automation of the Crozzle", *Australian Computer Journal*, 25, 2 , 1993, pp. 41-48.

[3] Rankin R.P., *Considerations for Rapidly Converging Genetic Algorithms Designed for Application to Problems with Expensive Evaluation Functions*, Ph.D. Dissertation, Dept. of Computer Science, University of Missouri-Rolla, 1993.