

Authors are encouraged to submit new papers to INFORMS journals by means of a style file template, which includes the journal title. However, use of a template does not certify that the paper has been accepted for publication in the named journal. INFORMS journal templates are for the exclusive purpose of submitting to an INFORMS journal and should not be used to distribute the papers in print or online or to submit the papers to another publication.

An OR approach for wORdle

Simran Lakhani

Indian Institute of Technology Bombay, simran@iitb.ac.in,

Mustafa Vora

Indian Institute of Technology Bombay, mustafa.vora@iitb.ac.in,

Ashutosh Mahajan

Indian Institute of Technology Bombay, amahajan@iitb.ac.in,

Abstract: Modeling exercises are designed around the popular game Wordle. These exercises can possibly be used in a computational or modeling laboratory course or in a course on modeling Operations Research problems. They assume some familiarity with modeling integer programs, a programming language like Python for basic text processing, and optimization software like AMPL or PYOMO. These exercises were tried in a masters-level course on modeling OR problems. Two sessions of three hours each are ideal for the whole exercise.

Key words: teaching modeling, wordle

Wordle is a popular single-player word guessing game (Victor (2022)). A player gets six chances to guess a five-letter target word. After every guess, the player receives feedback for each letter in the word as follows:

- i) If the letter is at the same position as in the target word, it appears green (G).
- ii) If the letter is present in the target word, but is not in the correct position it appears yellow (Y).
- iii) If the letter does not appear in the target word, it appears black (B).

Based on the feedback provided after each guess, the player can make better subsequent guesses. If the player correctly guesses the word in six trials, they win and lose otherwise. The gameplay is similar to the two-player board game Mastermind, which became popular during 1975 (Knuth (1976)). The game received a lot of interest from the OR community as well, and several studies were published to win it in the fewest possible attempts (Neuwirth (1982), Bernier et al. (1996), Kooi (2005)). Wordle is very similar to Mastermind except in Wordle, the guess should fit the

pattern and also be a valid word. Wordle has a dictionary of words known to the player, and the target word is selected from a subset of this dictionary. Several internet-based versions ([The New York Times Company \(2022\)](#), [Wordlegame \(2022\)](#)) of the game use words from the English dictionary. Following are a few examples of the color string that Wordle would show for the guess words.

- i) Target: POWER, Guess: ROPED. The feedback string is YGYGB.
- ii) Target: SMILE, Guess: SPREE. The feedback string is GBBBG. The first ‘E’ is black in the result and the second ‘E’ is at the correct position.
- iii) Target: AMPLE, Guess: KAPPA. The feedback string is BYGBB. Note that the guess word contains ‘A’ twice. The first instance is yellow, whereas the second is black, and similar is the case with ‘P’.

An OR Approach

One possible solution approach is to break the problem into two tasks: (a) first discover which letters appear in the target, and (b) then guess the target by rearranging the discovered letters (and some additional ones if needed). The first task can be modeled as a decision problem. The exercises below can be used as a guide to develop a mathematical model. The second task is somewhat easier and is left to the player. The resulting model is not perfect. It leaves some, hopefully easier, tasks to the player and does not even guarantee finding all letters. So this approach is just a rough tool to aid the decision maker and not a ‘solver’ for the problem. On the other hand, this approach is quite straightforward. The strategy is ‘fixed’ and independent of the feedback received in the first four tries. The player has to merely remember four simple words that can be used for every game. There are other interesting strategies ([Bertsimas and Paskov \(2022\)](#), [Anderson and Meyer \(2022\)](#)) that can be used to fully solve any given 5-letter Wordle but, not surprisingly, they are more complicated.

The goal of the following exercises is to find a fixed strategy so that the player can maximize the number of identified letters in the first four tries so that they can make “good” guesses in the remaining two attempts. Since there are 26 letters in the English alphabet, one needs six 5-letter words to discover all letters in the target. Even if one uses 6 such words, the letters appearing multiple times (for example, EAGLE) will not be known. Instead, one can try to pick four 5-letter words that cover 20 distinct letters. For example, the set $S_1 := \{\text{CLING, JUMPY, SHARK, TOWED}\}$ can be used to cover 20 letters, but it misses the letter ‘V’ that is more common than ‘J’. The exercises below attempt to find the best such four words.

Exercise 1: Can you find another set S_2 of 4 words that covers 20 distinct letters?

Exercise 2: Count the frequency of each of the 26 English letters in the dictionary. One can use the relevant parts¹ of Google Web Trillion Word Corpus (Norvig (2009b)) for a dictionary.

Exercise 3: Based on the above frequencies, can you decide whether S_1 is better than S_2 for the strategy described above?

Exercise 4: The dictionary has words with repeated letters, which should not be used for the above-described strategy. Also, one would like to avoid the six least frequent letters. Based on these two arguments, can you remove some words that will not be in an “optimal” strategy? How long is the filtered list after removing all such words?

Exercise 5: Which four words from the filtered list obtained above should you select so that all the 20 most frequent letters that you have identified appear exactly once?

One way to find such words is by formulating an integer programming model with a decision variable, say x_i that takes value one if word i from the filtered list is selected in the solution, and zero otherwise. Constraints are defined in a way to ensure that each of the 20 letters appears exactly once. Hence there can be twenty constraints each corresponding to a frequent letter. Report a solution after solving this model. It is likely the words you find are esoteric, proper nouns, or even abbreviations.

Exercise 6: Observe that the above optimization problem had no objective function. So can you introduce an objective to help get a better solution, that is easy to remember?

One objective could be to select those words in a solution that are more frequent and commonly used. The above dictionary of words also contains the frequency of how often each word is used. This frequency is obtained based on the occurrence of words in all the publicly available web pages (Norvig (2009a)). One way to do this exercise is by introducing an objective function that maximizes the total frequency of words selected. Can you model the complete optimization problem with this as an objective and constraints as described above? Report the solution and discuss your observations based on it.

The “optimal” solution obtained from above may have 1-2 commonly used words while others maybe not. Note that the frequency of the most frequent word ‘ABOUT’ in the filtered list is 1.2×10^9 , and that of the second most frequent word is 9.8×10^8 . A scatter plot of their frequencies (after sorting) shows an exponential decrease in frequency from ‘ABOUT’ to ‘IGOLE’ (the least

¹ A list is available online at <https://github.com/ashutoshmahajan/wordle-dict>

frequent word). Can you now explain why the optimal solution has rather uncommon words?

Exercise 7: One possible way to better model the objective is to take the logarithm of the frequency of each word. Since the logarithm grows slowly, the range from the highest frequent word to the lowest frequent word is not very drastic. It may keep the objective coefficients for all the variables within an order of magnitude only. Check if the new objective yields an easier-to-remember solution set.

Play the game with your friends or online and see if the above set is useful in solving the problem correctly. Can you think of a “hardest” target word where the above set provides minimum information (say only 1-2 letters)?

Acknowledgments

References

- Anderson BJ, Meyer JG (2022) Finding the optimal human strategy for wordle using maximum correct letter probabilities and reinforcement learning. *arXiv preprint arXiv:2202.00557* .
- Bernier JL, Herráiz CI, Merelo J, Olmeda S, Prieto A (1996) Solving mastermind using gas and simulated annealing: a case of dynamic constraint optimization. *International Conference on Parallel Problem Solving from Nature*, 553–563 (Springer).
- Bertsimas D, Paskov A (2022) An exact and interpretable solution to wordle. Preprint on webpage <https://mitsloan.mit.edu/ideas-made-to-matter/how-algorithm-solves-wordle>.
- Knuth DE (1976) The computer as master mind. *Journal of Recreational Mathematics* 9(1):1–6.
- Kooi B (2005) Yet another mastermind strategy. *ICGA Journal* 28(1):13–20.
- Neuwirth E (1982) Some strategies for mastermind. *Zeitschrift für Operations Research* 26(1):B257–B278.
- Norvig P (2009a) Natural language corpus data. Segaran T, Hammerbacher J, eds., *Beautiful data: the stories behind elegant data solutions*, 219–242 (“O’Reilly Media, Inc.”).
- Norvig P (2009b) Natural language corpus data: Beautiful data. URL <http://norvig.com/ngrams/>.
- The New York Times Company (2022) URL <https://www.nytimes.com/games/wordle/index.html>.
- Victor D (2022) Wordle is a love story. URL <https://www.nytimes.com/2022/01/03/technology/wordle-word-game-creator.html>.
- Wordlegame (2022) URL <https://wordlegame.org/>.