A Little Wordle Play

Robert Petit

March 2022

Introduction

The first guess for a WordleWord is an interesting microcosm of arbitrary, hyper-optimizable strategy decisions. One of the interesting catches of Wordle as a whole is trying to work through the results that you obtain throughout your guesses. The first word is unique: much like a chess opening, the game hasn't shaped up. Here, we will attempt to provide a small selection of "best first words" for a number of different strategies:

- Letter Hunting
- Position Hunting
- Letter & Position Combo
- Vowel Minimization

This provides you with the choice: do you want to get as many yellow letters or green letters as possible. We can weight and combine these scores to get a ranking that compromises between yellow and green results as well. Since vowels are fairly trivial to include on later guesses, we may be more interested in maximizing our correctly-guessed consonants and ignoring the contribution of vowels to our score.

On Judgement

In order to do this, we are going to use two different sets of words to treat as our "potential guesses" for the game. We have a large list (8497 words) of all 5 letter words in the English language. This is *technically* the full breadth of possible words in Wordle, but it includes words like "aalii" (a type of bush found in Hawaii) and "ganch" (to execute via impaling) in the list. While these are possible, they are not particularly likely to show up. To account for this, we also have a list of commonly used words (5757. We will use both the full short word lists to judge our results.

This will yield a set of results for each strategy: results when our scope is *all* words and results when our scope is *likely* words. This is four combinations of possible results:

- All words judged on all words
- All words judged on likely words
- Likely words judged on likely words
- Likely words judged on all words

We are only concerned with the first three. The last of these is useless, after all, there is no world in which you only want to *guess* a common word, but a possible result is *any* word.

Letter Hunting

We make the simple observation that

$$P(1\{x \in W\}) = E[1\{x \in W\}] = \frac{1}{n} \sum_{i=1}^{n} 1\{x \in w_i\}$$

Where x is a given letter, W is a random word and w_i is a specific word within our target sample. We then calculate $P(1\{x \in W\})$ for each unique letter in a word a form the expected yellow results, Y as

$$Y = \sum_{j=1}^{u} P(1\{x_j \in W\})$$

where the subscript j denotes a specific letter and u is the number of unique letters in that word. This gives us the expected number of yellow results any one word should have; naturally, we simply want this value to be the highest possible.

Position Hunting

This is largely similiar to letter hunting, but instead of being concerned with $P(1\{x \in W\})$ we now need to calculate the letter being in the correct spot, which is given by

$$P(1\{x_j = W_j\}) = E[(1\{x_j = W_j\})] = \frac{1}{n} \sum_{i=1}^{n} 1\{x_j \in w_{ij}\}$$

We then calculate the expected number of green results, G, as

$$G = \sum_{j=1}^{5} P(1\{x_j = w_{ij}\})$$

Combined Scoring & Vowel Omission

Once we have calculated a Y and G score for a word, we can combine these two scores based on a given weight t to give us the combined score C

$$C = tG + (1 - t)Y$$

Of note with the C score is to remember that we do discredit green results when searching for yellow results. A yellow letter and a green letter will both count for our purposes as a "yellow outcome", but only a green letter will count as a "green outcome". When we look at our full score, it's important to remember that the trade-off is between having results hit in general and having a specifically green result result instead. Each of yellow letters has $P(Green) = 0.2$, so there is a *very* strong relationship between them. This becomes consequential when we are choosing our weighting factor t, since at low values of t we would expect remarkable ranking similarity between different choices.

Finally, we can reconsider the Y, G, and C values but ignore the vowels when calculating each score in order to get a separate vowel-omitted score.

Results

The fun stuff!

Letter-Seeking

This is our Yellow-Maxing list, seperated by scope of the guesses and scope of the possible words.

Full List judged on Full List

Word	Y	Y (Consonants)	Word	Y	Y (Consonants)	Word	Y	Y (Consonants)
arose	1.773	0.596	arose	1.909	0.777	arose	1.909	0.777
orate	1.771	0.594	raise	1.884	0.777	arise	1.884	0.777
arise	1.769	0.596	arise	1.884	0.777	raise	1.884	0.777
raise	1.769	0.596	tears	1.871	1.031	serai	1.884	0.777
serai	1.769	0.596	rates	1.871	1.031	aster	1.871	1.031

Short List judge on Short List

Word	Y	Y (Consonants)
arose	1.909	0.777
raise	1.884	0.777
arise	1.884	0.777
tears	1.871	1.031
rates	1.871	1.031

Full List judged on Short List

Word	Y	Y (Consonants)
arose	1.909	0.777
arise	1.884	0.777
raise	1.884	0.777
serai	1.884	0.777
aster	1.871	1.031

Position Seeking

Green-Maxing yields the following sets:

Full List judged on Full List

Word	G	G (Consonants)
soree	0.651	0.228
salay	0.607	0.339
boree	0.592	0.169
sairy	0.586	0.338
saily	0.586	0.338

Short List judge on Short List

Word	G	G (Consonants)
sores	0.886	0.514
sales	0.874	0.499
soles	0.870	0.499
sates	0.855	0.480
sires	0.844	0.514

Full List judged on Short List

Word	G	G (Consonants)
soles	0.870	0.499
cones	0.825	0.454
sures	0.820	0.514
tales	0.813	0.439
mores	0.812	0.440