International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056
Volume: 09 Issue: 06 | June 2022 www.irjet.net p-ISSN: 2395-0072

Optimizing Wordle Helper using Datasets and Commonly used words.

Sarosh Dandoti[1]

Abstract - *Wordle is a very simple game that went viral at the start of 2022 and yet the game is very difficult without sufficient knowledge of vocabulary. Such simple 5-minute brain teasers like wordle, hangman look very simple at sight, but there is a deeper understanding of these games and figuring out if there's a statistical and informed approach one can make to maximize the win states. The game looks very random at the start, there are methods and initial guesses which eliminate doubts and restrict your search.*

Key Words: Wordle, Python, Programming, Logic, Datasets, Brain-Teasers.

1. INTRODUCTION

In Wordle, the player has 6 chances to guess a 5 letter word. These configurations can be changed as required hence changing the difficulty of the entire game. We have implemented an AI which suggests the best suitable word to maximize our win and get it at an earlier stage. The player gets more points if the correct word is guessed at a second chance than the last chance. You have to enter a letter word, and if the letters in that word lie in the goal word they turn yellow, if they turn green, it means that the position and word are correct, and if it turns grey it means that letter is not present in the goal word. This looks pretty simple when you play it yet one gets stuck due to the lack of vocabulary or simple the pressure of having fewer chances each time you play. To make this even more pressuring, the original wordle gives only 1 word per day. It also has a hard mode, which then forces you to use those letters which were green again on the position for the next guess, for example, if you hit the green letter 'A' on the first position, you have to guess all the next words starting with 'A'. similarly, if you hit let's say, 'W' in the middle, you have to use 'A' and 'W' both and then think of a word. This makes the game very difficult but the solution is guaranteed in lesser steps as there will be very few words matching those criteria. There may be multiple same letters to confuse the player. For example, if the goal word is 'sweet' and if the player hits 'e' at one position he may never try 'e' again assuming that the letter does not occur again in the words and hence losses the game.

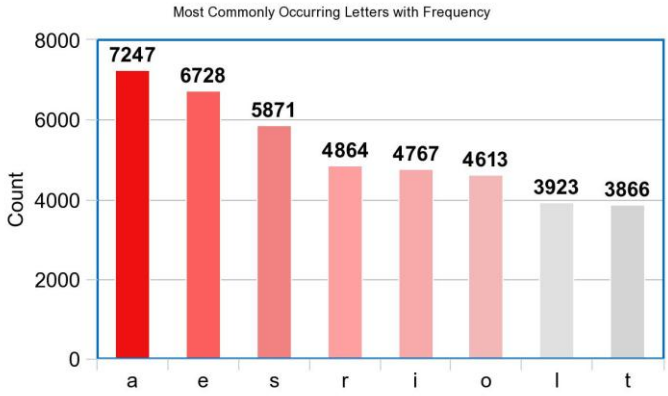
2.1 Data

To create our Wordle Helper we need lots of data containing 5 letter words. The dataset is an oxford dictionary words dataset from which only 5 letter words were selected. The data should be accurate and should not contain spelling mistakes. These words range from simple words like 'hello' to unique words like 'vraic'. This is an interesting part where one must find out which words are important and which ones are not. The Data must be converted into lowercase letters as computers consider 'A' and 'a' with unique ASCII values. Words that are very unique like mentioned above are sometimes not even in the Wordle Dictionary because, the dictionary is made up of commonly used words and since the player does not know the dictionary of words, it's difficult to find a word that fits appropriately as there may be multiple words that may fit into certain criteria. For resolving this we need to find the most commonly used words in the day-to-day language and the letter which are most occur in our 5 letter words dataset.

Here is a word cloud showing the most unique and least used words from the dataset.

These words even though are correct and identified by all the dictionaries in the world, are not used in day-to-day conversations and hence not recognized by the wordle dictionary.

And here are just s few most commonly occurring words in the dataset.



2.2 The Helper AI

This AI helper is made of two components. The first component is finding the most common word in the dataset. For a further explanation, we will refer to the words dataset as a wordlist. The Helper goes through the entire wordlist and finds the most commonly occurring letters. Let's Consider the first 5 most common letters. The Helper tries to find a 5 letter word using the combination of those 5 selected letters. If it cannot find a suitable word from our wordlist, it will also consider the 6th most commonly occurring letter. Here we can also add an additional feature where, if the Helper needs to consider the 6 letter, it will remove the letter with the lowest frequency from the first 5 letters. Doing this does increase performance but there may be some words that may disappear. From the wordlist the most commonly 5 occurring words were, 'a' , 'r' ,' i' ,'s' ,'e' .We can clearly see there can be a few words created with these letters like 'arise' and 'raise'. The best word to choose among these would be 'arise'. The reason we chose this is the work of the second component of this Helper, which suggests the best words from the given letters.

The second component basically uses the frequency of these occurring letters to determine which word would be best suitable and maximizes winning.

Like the above example, one could choose 'arise' or 'raise',

'arise' is more suitable because there are more words starting with the letter 'a than 'r'. This makes our guessing chances more accurate. This is the list of most commonly occurring words with their word frequency from our data.

({**'a'**: 7247, **'e'**: 6728, **'s'**: 5871, **'r'**: 4864, **'i'**: 4767, **'o'**: 4613, **'l'**: 3923, **'t'**: 3866, **'n'**: 3773, **'u'**: 3241, **'d'**: 2639, **'c'**: 2588, **'y'**: 2476, **'m'**: 2361, **'h'**: 2223, **'p'**: 2148, **'b'**: 1936, **'g'**: 1867, **'k'**: 1663, **'w'**: 1160, **'f'**: 1115, **'v'**: 853, **'z'**: 435, **'j'**: 372, **'x'**: 357, **'q'**: 139 })

There are 5 words that can be formed with the top 5 letters.

['aesir', 'aries', 'arise', 'raise', 'serai']

But since wordle works on the commonly heard and used words using 'arise' and 'raise' makes more sense. A strategy players use and which can be used by this Helper as well is to use words with the most commonly occurring letters and no repetition of the letters, So to eliminate options and shorten our search list one can use completely different words w.r.t alphabets as the second word, For example,

1. First Word: ARISE
2. Second Word: CLOUT
3. Third Word: NYMPH

The idea here is to eliminate as many words as possible so one can shorten the search list and get the guaranteed final answer on the fourth try. This method almost always guarantees the answer but it takes about 3 tries to work it out. This will be further explained with a proper example.

The other strategy is to go traditionally. This method works like the hard mode of the game.

2.3 GAME PLAYING

Here the Helper first suggests a word like 'arise' to start with. Notice that the word 'arise' has 3 vowels, hence also the reason why it helps us shorten our wordlist. The helper now matches the words with arise and through wordle, we now know the status of each letter, green, yellow, or grey. We input this information back into the Helper so that it can run through the list and give us a shorter list. If a certain letter is in the goal word but not in the correct position (yellow) , the helper will avoid the words with that letter in that position. Similarly, on each turn, the helper will shorten the list of words. And among those words, the AI finds the most likely word using the

method explained before. The parameters the AI Helper considers on each turn:

contains = [ ]

- enter the yellow words here

contains_pos = [ ]

- enter the position of the yellow words
-

fixed = [ ]

- enter the green words here

fixed_pos = [ ]

- enter the position of the green words

exclude = [ ]

- enter the grey words

This allows it to shorten the list and suggest the words fitting in the category.

3. EXAMPLES

We have discussed two methods to approach this problem Lets us discuss them with proper examples.

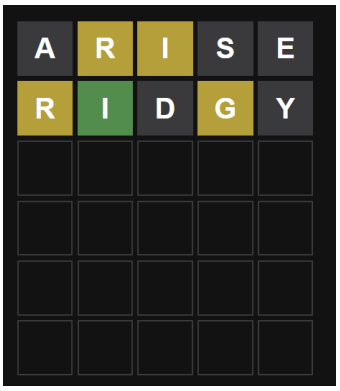
1. The Hard Mode Approach

This gives more words but a chance of getting early success.

Now select a word using these parameters. The words will obviously be many so we guess it based on the frequency as explained above.

So for the next word we skip the top letters 'a,e,s', So among the top occurring letters, we can now guess a word starting from 'r'. After cutting and shortlisting we have words like
wordlist:
[ricin','ridgy','rifty','right','rigid','rigol','rigor','robin']

'ridgy' would be the word to choose since we are following that frequency order letter by letter through the words.



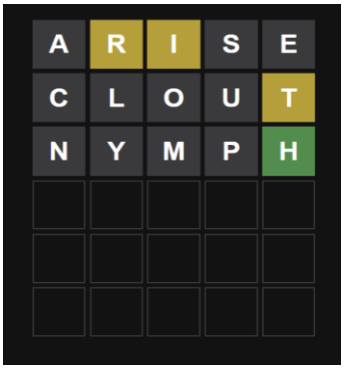
Do the same procedure again and get a shortlist of words then find the most suitable word.

wordlist : ['giron','girth','pirog','vigor']

Here you can try the word which seems more suitable.
But in the end, there may be a few cases where some words which under even those granular conditions.

2. The Elimination Method

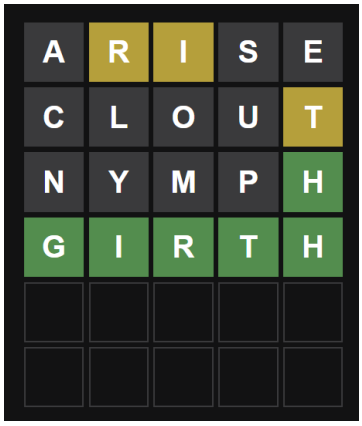
As said before this method tries to eliminate letters first as much as possible, so the solution will almost always hit on the 4th try.



Now we have available words as:

['birth','firth','girth']

Here is where sometimes we hit a problem. All the words here are correct and can be used in the wordle. But we don't know which is the right one. Even though we can try the frequency trick, since it's the only word here there is really a 33% chance either one is right here it's a risk one has to take. But such is the game of Wordle. Since there are so many words to deal with there will be cases where it will be an equal probability that either option is right. No Artificial Intelligence can make a correct guess without the extra knowledge of the problem. Here the solution can be anything of the three words.



4. Future

The work on this problem is continuously being worked on and trying to find a new and more optimized way of playing the game. The Helper is continuously being worked on to improve and have a correct dictionary of words to help in creating the suggestions. A new technique is being studied to predict these words by comparing their usage in day-to-day text conversations vs literature words.

5. Conclusion

This program works correctly and is tested. Wordle is an interesting puzzle where no matter how complicated a system is, at least some part of the final solution rests on probability and guessing. In this paper , we have discussed two approaches to the Wordle Game.

REFERENCES

[1] Finding the optimal human strategy for Wordle using maximum correct letter probabilities and reinforcement learning. Benton J. Anderson, Jesse G. Meyer
doi.org/10.48550/arXiv.2202.00557
1 Feb 2022

[2] Using Wordle for Learning to Design and Compare Strategies. Chao-Lin Liu.
doi.org/10.48550/arXiv.2205.11225
30 Apr 2022

[3] Finding a Winning Strategy for Wordle is NP-complete. Will Rosenbaum.
doi.org/10.48550/arXiv.2204.04104
8 Apr 2022