

Using information entropy to 'solve' Wordle.

Whilst exploring options under my favourite mathematical topic, I stumbled upon a brilliant YouTube video, which briefly explained how probability and frequency provided the most consistent and efficient approach to the word game Wordle. Wordle is a five-letter word guessing game and the player has 6 total guesses to win. On average, a regular Wordle player guesses the daily word in about four attempts, with three or fewer guesses being considered impressive. What makes the game easier is that the players receive information, based on their previous guess, about the correct word. Although the maths discussed in this essay can improve the average guess count, it does not guarantee beating a human. In other words, there is some amount of unavoidable luck involved. On the wordle website, there is a displayed list of 2,315 correct answers. These are part of a second list of 12,972 allowed guesses. These values will be used in the rest of the essay.

An example game is displayed below:

C	R	A	N	E
S	H	O	U	T
H	U	R	R	Y

The three colours: amber, green, and grey all have their separate meanings. Grey means neither the letter nor its placement are correct. Amber means the letter is correct, but it is in the wrong place. Hence, Green means both the letter and its placement are correct. The colours of the letters in a word will be described as a 'colour-set'.

The initial guess of CRANE gives us 'information' about its five letters in relation to the correct word HURRY. The R is amber; therefore, I know it is correct but in the wrong place. The other four letters (CANE) I know are not in the word. This follows through other guesses.

Introduction to Information theory:

The first part to 'solving' Wordle is to understand Information theory. Information theory is the quantification of the amount information gained after an event or, in this case, the outcome of a guess in Wordle (the colours of the letters). At first glance, information may seem to be more of a qualitative value. However, during World War Two, Claude Shannon, through his research into telecommunications, first devised how to quantify information. Information theory is based upon the probability of that event or colour-set occurring. If an event that occurs is surprising, therefore having

a lower probability, then we would gain more information and the opposite happens when an event occurs that is more common (higher probability). This displays an inverse relationship between probability of a guess and the quantity of information it provides.

Example for clarity: Deck of 52 cards

Imagine this scenario: A friend chooses a random card from a deck and asks you to try and guess it in as few tries as possible. He helps you by giving you information (card suit or number) about this card based on your previous guess (similar to the Wordle model). In this case, you guess a card that is a diamond. Your friend tells you that you have guessed the correct suit. This provides more information, for your next guess, than if you had guessed the wrong suit. This is because the number of probable options for this card is cut by $\frac{3}{4}$ rather than $\frac{1}{4}$. Looking at the probabilities of these situations:

Guessing the correct suit follows:

$$P(\text{Correct suit}[\text{diamonds}]) = 0.25$$

Whereas the wrong suit:

$$P(\text{Wrong suit}[\text{hearts, spades and clubs}]) = 0.75$$

As I have already mentioned, the probability of an actuality is inversely related to the information it provides. We now need to quantify this information. The unit of information I will be using is bits (bear in mind that other units can be used). A bit can be explained using the same example. When we correctly guessed the suit of the random card, we cut the space of possible cards down by a factor of 4 (where $p(x) = 0.25$). This is equal to 2 bits of information. If it was cut by a factor of 8 ($p=0.125$) it would be 3 bits, and so on... You can see the pattern here with powers of 2. This leads us to the equation for information:

$$h(x) = \log_2 \frac{1}{p(x)}$$

Where $h(x)$ denotes the information gained for a random variable x and where $p(x)$ is its probability.

The equation shows $1/p(x)$, which is in line with the inverse relationship we discovered earlier. This also ensures there are no negative values for $h(x)$. This equation can be rearranged to:

$$h(x) = -\log_2 p(x)$$

Applying this to our Wordle example from earlier, with our initial guess of CRANE:



The outcome of the guess CRANE, where the R is amber and all other letters are grey, leaves a space of possibility of 90 words out of the initial 2315 possible Wordle answers. This cuts the space of possibility down by roughly $1/25.72$, which is also the probability of the word containing an R and not C, A, N or E. Taking the log base 2 of 25.72, (the negative log of $1/25.72$) we get 4.68 bits of information. Therefore, from this guess and its outcome we deduce that we have gained 4.68 bits of information.

This is the basis of how information theory can be used to strategically solve Wordle. As I mentioned earlier, different base numbers can be used. Two, however, is the cleanest with its relation to how it cuts the space of possibility. Another common unit used is the nat, which is the natural unit for information. It uses the log base of e (Euler's number).

Information entropy:

Working out how much information a word gives us, after it has been guessed does not give us any strategic value in choosing what word we should have guessed. This leads on to the next part of information theory, which is information entropy. Information entropy is the expected amount of information a guess will yield. This is crucial when differentiating between what words to guess.

The information entropy of a guess can be worked out by taking the sum of every single one of its possible colour-set outcome's entropy. Firstly, to find a colour-set's singular entropy, one must calculate it's individual information content (in bits) and then multiply it by the probability of it happening (different colour-set outcomes will most likely have different probabilities). Then one must add all the colour-set's singular entropies together. This cannot be done on paper and computers should be used, especially when comparing hundreds of different words' entropies.

Let us consider a new example, where we apply entropy to the same word CRANE, but it shows a different colour-set outcome.



The guess CRANE informs us that the correct answer contains an R, A and N but not in the shown positions and it also does not contain a C or E.

After using a computer to select all the words with the highest expected information content (entropy), it comes up with RAYON.

Explaining the information entropy of RAYON:

RAYON possible colourings:

R	A	Y	O	N	R	A	Y	O	N	R	A	Y	O	N	R	A	Y	O	N	R	A	Y	O	N
R	A	Y	O	N	R	A	Y	O	N	R	A	Y	O	N	R	A	Y	O	N	R	A	Y	O	N

As shown above, there are ten colour-sets that are possible after guessing the word RAYON, and fortunately they all lead to a single unique answer. This means that each colour-set has a $1/10$ probability of occurring. This means that they each have an information content of $-\log_2 \frac{1}{10}$ which roughly equals 3.32. If we multiply this information quantity by their likelihood of happening (again $1/10$) and then add all of the colourings together, we then get an expected information content of 3.32, which is, according to a computer programme (link below), the highest out of any of the allowed guess list (12972 words).

This equation is shown below:

$$H(x) = - \sum_x p(x) \cdot \log_2 p(x)$$

Where $H(x)$ denotes the expected average information or information entropy.

Conclusion:

To finalise, applying this maths to 'beat' Wordle requires quite bit of computing work, which I am unfortunately unable to do. A computer should analyse the expected information of every allowed word in the Wordle dictionary and pick out the word which gives the highest average information content (entropy). For the first word, the space of possibility is just the set of Wordle possible answers (2315 words). After receiving information on that first guess, whether it be more or less than what was expected, there will be a new set of possible answers. A second word, again with the highest information entropy, should then be guessed. This should be continued until the word has been reached. The best starting word, which leads to the lowest average guess count (the word with the highest information entropy), is SALET. This is closely followed by our example CRANE and other words such as TRACE. An example of a website that solves Wordle through information entropy has been made by Jonathon Olson (link below). It has compiled results for a high number of trials and has found that the computers average guess count is around 3.42. I suggest you have a look since it is fascinating.

How does this compare with average human Wordle statistics?

A study done by Matic Broz, showed that, globally, the average number of guesses is 4.016, with Sweden having the lowest national average with 3.72 and the UK national average is 3.89. This displays how a computer following information theory throughout (average guess of 3.42) beats the global average by some margin and is still some distance below the smart Swedes. The question is can you beat information entropy.

Cameron Abbasian

Bibliography:

- 3Blue1Brown YouTube video - <https://www.youtube.com/watch?v=v68zYyaEmEA&t=1309s>
- <https://machinelearningmastery.com/what-is-information-entropy/>
- Wordle allowed guess list and answer list - <https://gist.github.com/cfreshman/dec102adb5e60a8299857cbf78f6cf57>
- Computer Wordle solver website - <https://jonathanolson.net/wordle-solver/>
- Global human Wordle Statistics - [https://photutorial.com/wordle-statistics/#:~:text=On%20average%2C%20Wordle%20players%20can,the%20theoretical%20best%20\(0.043%25\).](https://photutorial.com/wordle-statistics/#:~:text=On%20average%2C%20Wordle%20players%20can,the%20theoretical%20best%20(0.043%25).)
- Map of national wordle averages - <https://www.forbes.com/sites/mattgardner1/2022/03/05/wordle-stats-reveal-10-best-countries-and-america-misses-out/>
- Claude Shannon's Information theory essay - <https://people.math.harvard.edu/~ctm/home/text/others/shannon/entropy/entropy.pdf>