Adam Livingston

4 May 2022

BASV 376 Mathematics for Applied Technology

Assistant Professor Emily Carroll

Practical Exercise 6

**Tasks**

1. *Compute the entropy of each text excerpt, using words as the unit of comparison.*
2. *Find the frequency of each word in each text.*
3. *Find the information content of each word in each text.*
4. *Find the entropy in each text.*
5. *Answer the questions in Quiz #6 to check your computations.*

***Student’s Methodology:***

1. Take excerpt from PDF and paste into Microsoft Word
2. Remove paragraphs to make one, large paragraph.
3. Remove erroneous symbols from text (extended hyphens, quotations, etc.)
   1. Ensure that hyphenated words are kept. Hyphenated words such as “bare-branched” are considered one word.
4. Paste text in one cell of Excel document.
5. Data > Text to columns > delimited
   1. This takes every word and places it in a separate column on one row
6. Copy row. Transpose into column.
7. Pivot table with wordcount as a column.
   1. Take note of **total** wordcount = n.
8. Create new columns for Probability, Information, and Entropy.
9. Assign equations for each column and calculate.
   1. Probability: “=frequencyCell / total wordcount”
   2. Information: “=((-LN(frequencyCell))/LN(2))”
   3. Entropy: “=probabilityCell\*informationCell”

**Questions**

1. *Which text has mathematically higher entropy, when words are treated as independent events?*

*The Great Gatsby* has mathematically higher entropy given the fact that even though there is a similar overall quantity of words in it as there are in *Farewell to Arms*, there are more instances of *unique* words. Even though the quantity of *unique* words does not directly influence the equations themselves when calculating entropy, it has a direct correlation to the entropy. The ratio of unique-to-total words would indicate that if there is a higher ratio, then there is a higher amount of entropy.

Taking Example 2 from the lecture when determining the weather forecast in Sierra Vista, AZ, if a different, unique weather event occurred every day, then the ratio of *unique* weather events to *total* weather events 1:1. The probability of each individual weather event would be 0.002. The information of each event would be 8.9 bits. This would make the average information content in a weather forecast (aka entropy) would be 0.002\*8.9\*365 = 6.5 bits, an 8.5x higher average information content than the previously-calculated 0.764 bits

In the test that was given above of a 1:1 ratio of unique-to-total, the only influencing factor that determines the entropy would be n total events. If there were 730 days in one year (twice as many) with a 1:1 ratio of unique-to-total weather events, then the entropy would be 13 bits.

1. *Is it a fair assumption to treat words as independent events?*

No. When mathematically calculating the information content in a piece of writing and treating words as independent events, the information content of a piece is higher based on the ratio of unique-to-total words. Also, words are dependent events since sentence structure, syntax, grammar, etc. are influenced on the type of word(s) that is/are preceding it. Treating words as independent events means that the order of words doesn’t matter.

1. *Which author do you think conveys more information per word in their writing? Does this match your answer for question 2a? Why do you think that is? (there is not one correct answer to this question).*

Mathematically, F. Scott Fitzgerald conveys more information-per-word in their writing at a rate of 7.56 bits/word versus Hemmingway at 6.7 bits/word as determined in question 2a.

However, humans’ interpretation of “information” is subjective. An author can mathematically convey a high amount of information by writing gibberish as long as every word is unique as explained in question 1 with the weather example. The usefulness of words and writing cannot be stated in terms of mathematical “information”. It would have to be placed in a different category of audience “value”. *Modern Quantum Mechanics* conveys lots of “information”, but would have no “value” when read to a toddler (or adults, for that matter) in comparison to *Green Eggs and Ham* or *Cat in the Hat*.

A mathematical example of this concept is if an author wrote every word in the dictionary in random order without definitions. This can mathematically convey a high amount of information that has little-to-no value to the audience. With about 170k words in current use:

p(m) = 1/170,000

h(m): 17.38 bits

Entropy: 2,953,779.8 bits

Value: None, besides spelling reference but no efficient way to access that information.