

HW 5: Reading Questions and Exercises

Due: Monday, October 29

Submit *short* computer-formatted, spell-checked answers to the following (one or two paragraphs per question).

1. What reason does Braitenberg give for asserting that Vehicle 12 has free will? Do you agree with him?
2. Using Google, estimate the probability of each bigram (pair of adjacent words) in the following two sentences: “Genius is born, not made”, and “Genius is born, not paid”. Use these values to estimate the probability of each sentence.
3. Repeat the exercise above, but use trigram probabilities instead of bigram probabilities.
4. What is the purpose of “smoothing” in n-gram analysis? Give an example of a next-word-prediction task in which smoothing would be important.
5. In class, we saw that the method of Good-Turing discounting revises n-gram counts according to the following formula:

$$c^* = (c + 1) \frac{n_{c+1}}{n_c},$$

where n_c is the number of n-grams appearing c times in the data. Explain why this is a good formula to use.

6. Describe two possible weaknesses of the n-gram method for predicting the next word in a text or in speech.
7. Using the vectors given in the “fried chicken” retrieval example given in class, calculate the similarity of the query “fried chicken” and each of the two recipes.
8. Describe two possible weaknesses of the vector-space model for determining the relative similarity of two documents.