Natural Language Processing
CSCI 4152/6509 — Lecture 12
P0 Topics Discussion (4); N-gram Model (start)

Instructors: Vlado Keselj
Time and date: 16:05 – 17:25, 21-Oct-2024
Location: Carleton Tupper Building Theatre C

Previous Lecture

- Fully Independent Model (continued)
- Naïve Bayes classification model
  - Assumption, definition
  - Graphical representation
  - Spam detection example
  - Computational tasks
  - Number of parameters
  - pros and cons, additional notes
  - Bernoulli and Multinomial Naïve Bayes
- P0 Topics Discussion (3)

P0 Topics Discussion (4)

- Discussion of individual projects as proposed in P0 submissions
- Projects discussed: P-14, P-15, P-16, P-17, P-18, P-19, P-20, P-21, P-23, P-24

N-gram Model

- Let us first introduce *language modeling*
- *Language Modeling:* Estimating probability of arbitrary NL sentence: P(sentence)
- Example: Speech recognition

$$\begin{aligned}
\arg\max_{\text{sentence}} \mathrm{P}(\text{sentence}|\text{sound}) &= \arg\max_{\text{sentence}} \frac{\mathrm{P}(\text{sentence}, \text{sound})}{\mathrm{P}(\text{sound})} \\
&= \arg\max_{\text{sentence}} \mathrm{P}(\text{sentence}, \text{sound}) \\
&= \arg\max_{\text{sentence}} \mathrm{P}(\text{sound}|\text{sentence})\mathrm{P}(\text{sentence})
\end{aligned}$$

- Acoustic model and Language model

Language Modeling

- Task of estimating probability of arbitrary utterance in a language
- Alternative task: Predicting the next token in a sequence: e.g., the next word or words, in a sentence, or next character or characters
- N-gram model: a "natural" model for this task

N-gram Model Assumption

$$\mathrm{P}(w_1 w_2 \ldots w_n) = \mathrm{P}(w_1|\cdot\cdot)\mathrm{P}(w_2|w_1\cdot)\mathrm{P}(w_3|w_2 w_1)\ldots\mathrm{P}(w_n|w_{n-1}w_{n-2})$$

N-gram Model: Notes

- Reading: Chapter 4 of [JM]
- Use of log probabilities
  - similarly as in the Naïve Bayes model for text
- Graphical representation