Chapter 3:
Basics of Language Modeling
Section 3.1. Language Modeling in Automatic Speech Recognition (ASR)

All graphs in this section are from the book by Schuckat-Talamazzini unless indicated otherwise.
### Complexity of Speech Recognition Applications

<table>
<thead>
<tr>
<th>Feature</th>
<th>Complexity</th>
</tr>
</thead>
<tbody>
<tr>
<td>Alarmstopschalter</td>
<td>1</td>
</tr>
<tr>
<td>Menü-Steuerung (J/N)</td>
<td>2</td>
</tr>
<tr>
<td>Zahlen/Ziffern</td>
<td>10(\pm x)</td>
</tr>
<tr>
<td>Gerätebedienung</td>
<td>20–200</td>
</tr>
<tr>
<td>Auskunftsdialog</td>
<td>500–2000</td>
</tr>
<tr>
<td>Alltagssprache</td>
<td>8,000–20,000</td>
</tr>
<tr>
<td>Diktiermaschine</td>
<td>20,000–50,000</td>
</tr>
<tr>
<td>Deutsch ohne Fremdwörter</td>
<td>ca. 300,000</td>
</tr>
</tbody>
</table>
What makes Speech Recognition difficult

Guten Morgen, Herr Hauptkommissar Thanner. Gibt es irgendeines Neues im Fall “Verbmobil”?

Morgen, Thanner. Irgendwas Neues im Fall “Verbmobil”?

morgen thanner irgendswas neues im fall verbmobil

morgenthannerirgendswanesimfallverbmobil

moangtannairgnwasneuesimfallvermobial

moangthannairgnwasneuesimfallvermobial

moangthannairgnwasneuesimfallvermobial

moangthannairgnwasneuesimfallvermobial

moangthannairgnwasneuesimfallvermobial

der Text in “Schönschrift”

spontan gesprochene Sprache

Großschreibung? Satzzeichen?

kontinuierliche Sprache

Aussprachevarianten

artikulatorische Verschleifung

Störungen & Verzerrungen des akustischen Kanals

Überlagerung d. Fremdstimmen „Cocktailparty-Effekt“
Milestones in ASR

DARPA SUR
- digitale Filterung
- Lineare Vorhersage
- Baum-Welch-Algorithmus
- schnelle DFT
- Dynamic Zeitchen (DTW)

DARPA Speech & Natural Language
- HARPY
- HMM zur Spracherkennung
- Sprecherverifikation
- Baum-Welch-Algorithmus
- Sprachsynthese vom Text
- LPC-Metrik
- DTW-Einzelworterkenner

- IBM Tangora
- stochastic N-Gramme
- mel-Cepstrum
- Verbundworterkennung
- LBG-Algorithmus
- backpropagation-Algorithmus
- Mikrofonfelder

- SPHINX
- n beste Wortketten
- TDNN
- Phonotopische Karte

Introduction into ASR

Source: Steve Young: the HTK book
Basic Components of an ASR System

Speech Signal

Feature Extraction

Stream of feature vectors $A$

Acoustic Model
$P(A|W)$

Language Model
$P(W)$

Search

$\hat{W} = \arg\max [P(A|W) P(W)]$

Recognized word sequence $\hat{W}$
The Markov Assumption

Cutting history to M-1 words:

\[ P(w_i | w_1, w_2, \ldots, w_{i-1}) = P(w_i | w_{i-M+1}, \ldots, w_{i-1}) \]

Special cases:

**Trigram** \[ P(w_i | w_1, w_2, \ldots, w_{i-1}) \approx P(w_i | w_{i-2}, w_{i-1}) \]

**Bigram** \[ P(w_i | w_1, w_2, \ldots, w_{i-1}) \approx P(w_i | w_{i-1}) \]

**Unigram** \[ P(w_i | w_1, w_2, \ldots, w_{i-1}) \approx P(w_i) \]

“**Zero-gram”** \[ P(w_i | w_1, w_2, \ldots, w_{i-1}) \approx \frac{1}{W} \]

(uniform distribution)
Section 3.2. A Measure for LM Quality: Preplexity
Motivation

- Speech recognition can be computationally expensive
- For research and development: need a simple evaluation metric for fast turn-around times in experiments
Test your Language Model

What’s in your hometown newspaper ??
Test your Language Model

What’s in your hometown newspaper today
Test your Language Model

It’s raining cats and ???
Test your Language Model

It’s raining cats and dogs
Test your Language Model

President Bill ???
Test your Language Model

President Bill Gates
Definition of Perplexity

Let $w_1, w_2, \ldots w_N$ be an independent test corpus not used during training.

**Perplexity**

$$PP = P(w_1, w_2, \ldots w_N)^{-1/N}$$

Normalized probability of test corpus
Example: Zerogram Language Model

“Zerogram” \( P(w_i | w_1, w_2, \ldots, w_{i-1}) \approx \frac{1}{W} \)

\[
PP = P(w_1, w_2, \ldots, w_N)^{-1/N}
\]

\[
= \left( \prod_{i=1}^{N} \frac{1}{W} \right)^{-1/N}
\]

\[= W \]

Interpretation: perplexity is the “average” de-facto size of vocabulary
Alternate Formulation

Assumption:

Use an $M$-gram language model

History $h_i = w_{i-M+1} \ldots w_{i-1}$

Idea:

collapse all identical histories
Alternate Formulation

Example:

M=2

Corpus:

“to be or not to be”

$h_2=\text{“to”}$  $w_2=\text{“be”}$

$h_6=\text{“to”}$  $w_6=\text{“be”}$

\[\rightarrow\text{calculate } P(\text{“be”} | \text{“to”}) \text{ only once and scale it by 2} \]
Alternate Formulation

\[ PP = P(w_1, w_2, \ldots, w_N)^{-1/N} \]

\[ = \left( \prod_{i=1}^{N} P(w_i \mid h_i) \right)^{-1/N} \]

Use Bayes decomposition

Try to get rid of product

\[ = \exp \left\{ \log \left( \prod_{i=1}^{N} P(w_i \mid h_i) \right)^{-1/N} \right\} \]

\[ = \exp \left\{ - \frac{1}{N} \log \left( \prod_{i=1}^{N} P(w_i \mid h_i) \right) \right\} \]
Alternate Formulation

\[ = \exp \left( - \frac{1}{N} \log \left( \prod_{i=1}^{N} P(w_i | h_i) \right) \right) \]

\[ = \exp \left( - \frac{1}{N} \sum_{i=1}^{N} \log(P(w_i | h_i)) \right) \]

\[ = \exp \left( - \frac{1}{N} \sum_{w,h} N(w,h) \log(P(w|h)) \right) \quad \text{N}(w,h): \text{absolute frequency of sequence h,w on test corpus} \]

\[ = \exp \left( - \sum_{w,h} f(w,h) \log(P(w|h)) \right) \quad \text{f}(w,h): \text{relative frequency of sequence h,w on test corpus} \]
Alternate Formulation: Final Result

\[ PP = P(w_1 \ldots w_N)^{-1/N} \]

\[ = \exp \left( - \sum_{w,h} f(w, h) \log(P(w \mid h)) \right) \]
Alternate Formulation: Zerogram Example

“Zerogram” \( p(w_i \mid w_1, w_2, \ldots, w_{i-1}) \approx \frac{1}{W} \)

\[
PP = \exp \left( - \sum_{w,h} f(w,h) \log(P(w \mid h)) \right) = \exp \left( - \sum_{w,h} f(w,h) \log \left( \frac{1}{W} \right) \right) 
\]

\[
= \exp \left( - \log \left( \frac{1}{W} \right) \sum_{w,h} f(w,h) \right) = \exp(\log(W) \ast 1) = W
\]

Identical result
Perplexity and Error Rate

Perplexity is correlated to word error rate

Power law relation
Quality Measure: Mean Rank

- **Definition:**
  - Let \( w \) follow \( h \) in the test text
  - Sort all words after a given history according to \( p(w_i|h) \)
  - Determine position of correct word \( w \)
  - Average over all events in the test text
Alternative: Average Rank

<table>
<thead>
<tr>
<th>The</th>
<th>are</th>
<th>to</th>
<th>know</th>
<th>the</th>
<th>issues</th>
<th>necessary</th>
</tr>
</thead>
<tbody>
<tr>
<td>This</td>
<td>will</td>
<td>have</td>
<td>this</td>
<td>the</td>
<td>problems</td>
<td>date</td>
</tr>
<tr>
<td>One</td>
<td>the</td>
<td>understand</td>
<td>these</td>
<td>the</td>
<td></td>
<td>information</td>
</tr>
<tr>
<td>Two</td>
<td>would</td>
<td>do</td>
<td>any</td>
<td>a</td>
<td>problem</td>
<td>above</td>
</tr>
<tr>
<td>A</td>
<td>also</td>
<td>got</td>
<td>the</td>
<td>a</td>
<td></td>
<td>other</td>
</tr>
<tr>
<td>Three</td>
<td>do</td>
<td>use</td>
<td>a</td>
<td></td>
<td>problem</td>
<td>time</td>
</tr>
<tr>
<td>Please</td>
<td>need</td>
<td>provide</td>
<td>them</td>
<td>all</td>
<td></td>
<td>people</td>
</tr>
<tr>
<td>In</td>
<td>We</td>
<td>insert</td>
<td>all</td>
<td></td>
<td></td>
<td>operators</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>role</th>
<th>and</th>
<th>the</th>
<th>next</th>
<th>by</th>
<th>two</th>
<th>meeting</th>
<th>of</th>
</tr>
</thead>
<tbody>
<tr>
<td>thing</td>
<td>from</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>that</td>
<td>in</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>to</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>contact</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>are</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>parts</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>with</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>point</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>were</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>for</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>requiring</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>issues</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>to</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>at</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>weak</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>days</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>1836</th>
<th>1837</th>
<th>1838</th>
<th>1839</th>
<th>1840</th>
<th>1841</th>
<th>1842</th>
<th>1843</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Important
### Quality Measure: Mean Rank

<table>
<thead>
<tr>
<th>Measure</th>
<th>Correlation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Perplexity</td>
<td>0.955</td>
</tr>
<tr>
<td>Mean Rank</td>
<td>0.957</td>
</tr>
</tbody>
</table>
Quality Measure: worst Score

- Which fraction of the test text have a score worse than $S_c$
- Correlation: 0.919
Summary

• Speech recognition
  • Basic task
  • Language model as a component
• Perplexity:
  • Measures quality of language model