CS15-319 / 15-619
Cloud Computing

Recitation 14
November 26th, 2013
Announcements

• Encounter a general bug:
  – Post on Piazza
• Encounter a grading bug:
  – Post Privately on Piazza
• Don’t ask if my answer is correct
• Don’t post code on Piazza
• Search before posting
• Post feedback on OLI
Module to Read

- UNIT 5: Distributed Programming and Analytics Engines for the Cloud
  - Module 16: Introduction to Distributed Programming for the Cloud
  - Module 17: Distributed Analytics Engines for the Cloud: MapReduce
  - Module 18: Distributed Analytics Engines for the Cloud: Pregel
  - Module 19: Distributed Analytics Engines for the Cloud: GraphLab
Quiz 5

• Quiz 5 Due Next Tuesday
  – Tuesday 12/03/2013 11:59PM Pittsburgh
  – Late submissions are NOT accepted

• Timed
  – 180 minutes once started
  – Remember to click SUBMIT before the deadline
Construct an n-gram model of the corpus

- An n-gram is a phrase with n contiguous words
- For example a set of 1,2,3,4,5-grams with counts:

<table>
<thead>
<tr>
<th>#</th>
<th>Example</th>
<th>Count</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>this</td>
<td>1000</td>
</tr>
<tr>
<td>2</td>
<td>this is</td>
<td>500</td>
</tr>
<tr>
<td>3</td>
<td>this is a</td>
<td>125</td>
</tr>
<tr>
<td>4</td>
<td>this is a cloud</td>
<td>60</td>
</tr>
<tr>
<td>5</td>
<td>this is a cloud computing</td>
<td>20</td>
</tr>
</tbody>
</table>
This Week’s Goal

Example:

this

<table>
<thead>
<tr>
<th>Options</th>
<th>Count</th>
<th>Probability</th>
</tr>
</thead>
<tbody>
<tr>
<td>this was</td>
<td>150</td>
<td>0.15</td>
</tr>
<tr>
<td>this is</td>
<td>500</td>
<td>0.50</td>
</tr>
<tr>
<td>this day</td>
<td>250</td>
<td>0.25</td>
</tr>
<tr>
<td>this kiss</td>
<td>25</td>
<td>0.03</td>
</tr>
<tr>
<td>this boy</td>
<td>75</td>
<td>0.08</td>
</tr>
</tbody>
</table>

Options Count Probability
this is 500 0.50
this day 250 0.25
this was 150 0.15
this boy 75 0.08
this kiss 25 0.03
This Week’s Goal

Build a statistical language model that contains the probability of a word appearing after a phrase

<table>
<thead>
<tr>
<th>#</th>
<th>Example</th>
<th>Count</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>this</td>
<td>1000</td>
</tr>
<tr>
<td>2</td>
<td>this is</td>
<td>500</td>
</tr>
<tr>
<td>3</td>
<td>this is a</td>
<td>125</td>
</tr>
<tr>
<td>4</td>
<td>this is a cloud</td>
<td>60</td>
</tr>
<tr>
<td>5</td>
<td>this is a cloud computing</td>
<td>20</td>
</tr>
</tbody>
</table>
This Week’s Goal

Store and index the words and their probabilities to use in an application.

Map-Reduce Program

N-Gram & Probability

HBase

Master
Core: 1
Core: 2
Core: 3
Core: 4

Configure & Link

Front-End Server
Upcoming Deadlines

• Project 4:

  Project 4
  Input Text Predictor: Language Model and User Interface
  Language Model Generation
  Checkpoint
  11:59PM
  12/6/2013

• Unit 5:

  UNIT 5: Distributed Programming and Analytics Engines for the Cloud
  Module 16: Introduction to Distributed Programming for the Cloud
  Module 17: Distributed Analytics Engines for the Cloud: MapReduce
  Module 18: Distributed Analytics Engines for the Cloud: Pregel
  Module 19: Distributed Analytics Engines for the Cloud: GraphLab
  Quiz 5: Distributed Programming and Analytics Engines for the Cloud
  Checkpoint
  Available Now
  Due 12/3/13 11:59 PM