Bringing Speech Technologies to the World

IAP 2016: Technologies for Speech and Language Processing

Mandy Korpusik and Tuka Al Hanai
Tuesday 19th Jan 2016
Today

1. Acoustic Phonetics
2. Language Modeling
3. Acoustic Modeling
4. Building a Recognizer
5. **Bringing Speech Technologies to the World**
**Activity:** Generate your own startup idea!

1. What is the **problem**?
2. What is your proposed **solution**?
   a. (as it relates to speech and language technologies)
3. **Judging Criteria:**
   a. Innovativeness, scalability, feasibility, market competition, business model
4. You are pitching to investors (us).
5. Describe your idea here (20 minutes):
   a. [https://goo.gl/FawJnk](https://goo.gl/FawJnk)
   b. Team up (2-3 people)
   c. Limit to 3 slides
   d. Talk 3 minutes
Outline

1. **App-building Tools**
2. **Current NLP Research**
   a. Human Computation
   b. Sentiment analysis
   c. Machine translation
   d. Dialogue Systems
   e. Question answering
3. **In-class Activity**
Amazon

Alexa Fund: $100 million investment for startups!

Alexa
Build engaging voice experiences for your services and devices

Alexa Skills Kit (ASK) Developer Preview
A free SDK that lets you easily add new voice capabilities

Alexa Voice Service (AVS) Developer Preview
Bring voice capabilities to your connected device

The Alexa Fund
$100 million in investment to fuel voice technology innovation
## IBM Developer Cloud

<table>
<thead>
<tr>
<th>Language</th>
<th>Speech</th>
<th>Vision</th>
<th>Data Insights</th>
</tr>
</thead>
<tbody>
<tr>
<td>AlchemyLanguage</td>
<td>Concept Expansion</td>
<td>AlchemyVision</td>
<td>AlchemyData News</td>
</tr>
<tr>
<td>Concept Insights</td>
<td>Dialog</td>
<td>Visual Insights</td>
<td>Tradeoff Analytics</td>
</tr>
<tr>
<td>Document Conversion</td>
<td>Language Translation</td>
<td>Visual Recognition</td>
<td></td>
</tr>
<tr>
<td>Natural Language Classifier</td>
<td>Personality Insights</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Relationship Extraction</td>
<td>Retrieve and Rank</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Word Count returns the percentage of words per tone and the total number of words for each individual tone trait.

Click on any highlighted word to see suggested synonyms.

https://tone-analyzer-demo.mybluemix.net/
Personality Portrait

12760 words analyzed: Very Strong Analysis

Summary

You are shrewd and somewhat inconsiderate.

You are empathetic: you feel what others feel and are compassionate towards them. You are proud: you hold yourself in high regard, satisfied with who you are. And you are laid-back: you appreciate a relaxed pace in life.

Your choices are driven by a desire for connectedness.

You consider helping others to guide a large part of what you do: you think it is important to take care of the people around you. You are relatively unconcerned with tradition: you care more about making your own path than following what others have done.

How did we get this?

You are likely to_____

☐ Click on an ad

☐ Follow on social media

☐ Change careers

You are unlikely to_____

☒ Treat Yourself

☒ Reply on social media

☒ Buy eco-friendly

https://watson-pi-demo.mybluemix.net/
For more than twenty years past I have been paying special attention to the question of Health. While in England, I had to make my own arrangements for food and drink, and I can say, therefore, that my experience is quite reliable. I have arrived at certain definite conclusions from that experience, and I now set them down for the benefit of my readers.

As the familiar saying goes, "Prevention is better than cure." It is far easier and safer to prevent illness by the observance of the laws of health than to set about curing the illness which has been brought on by our own ignorance and carelessness. Hence it is the duty of all thoughtful men to understand aright the laws of health, and the object of the following pages is to give an account of these laws.
Dragon NaturallySpeaking for developers

Learn more about how to add speech recognition to in-house and commercial applications or workflows.

Learn More about Dragon for Developers

Top Imaging OCR Toolkit for Windows, Linux and Mac

OmniPage SDK has everything you need to add powerful imaging, recognition, conversion and PDF features to your applications.

More About OmniPage Capture SDK

Nuance Mobile Solutions for Developers

Speech-enable your mobile app within minutes with the Dragon Mobile SDK.

More About Nuance Mobile Developer SDK

Nuance Healthcare Solutions for Developers

The 360 | Development Platform allows you to embed medical voice and understanding functionality in your applications.

360 | Development Platform
PowerShare Innovation
Useful Links

- **Build an iOS app like Siri**

- **Google Speech API**
  - [https://www.npmjs.com/package/google-speech-api](https://www.npmjs.com/package/google-speech-api)

- **HTML5 Speech API**

- **IBM Watson Speech API**
Outline

1. App-building Tools

2. Current NLP Research
   a. Human Computation
   b. Sentiment analysis
   c. Machine translation
   d. Dialogue Systems
   e. Question answering

3. In-class Activity
Collective Intelligence

Work Distribution Made Easy

Mechanical Turk gives businesses and developers access to an on-demand, scalable workforce

- **Flexibility**: Scale your workforce up and down quickly
- **Accuracy**: Get high-quality, cost-effective results
- **Speed**: Start receiving results in minutes

Get Started
Video Annotation

http://web.mit.edu/vondrick/vatic/
Text Summarization

Automatic clustering generally helps separate different kinds of records that need to be edited differently, but it isn't perfect. Sometimes it creates more clusters than needed, because the differences in structure aren't important to the user's particular editing task. For example, if the user only needs to edit near the end of each line, then differences at the start of the line are largely irrelevant, and it isn't necessary to split based on those differences. Conversely, sometimes the clustering isn't fine enough, leaving heterogeneous clusters that must be edited one line at a time. One solution to this problem would be to let the user rearrange the clustering manually, perhaps using drag-and-drop to merge and split clusters. Clustering and selection generalization would also be improved by recognizing common text structure like URLs, filenames, email addresses, dates, times, etc.

Automatic clustering generally helps separate different kinds of records that need to be edited differently, but it isn't perfect. Sometimes it creates more clusters than needed, because the differences in structure aren't relevant to a specific task. Conversely, sometimes the clustering isn't fine enough, leaving heterogeneous clusters that must be edited one line at a time. One solution to this problem would be to let the user rearrange the clustering manually using drag-and-drop edits. Clustering and selection generalization would also be improved by recognizing common text structure like URLs, filenames, email addresses, dates, times, etc.

http://projects.csail.mit.edu/soylent/
Photo Annotation

Word Pronunciation: Quizlet
Outline

1. App-building Tools
2. Current NLP Research
   a. Human Computation
   b. **Sentiment analysis**
   c. Machine translation
   d. Dialogue Systems
   e. Question answering
3. In-class Activity
Sentiment Analysis: Positive or Negative?

This movie *doesn't* care about cleverness, wit or any other kind of intelligent humor.

There are slow and repetitive parts, *but* it has just enough spice to keep it interesting.
Challenges of Twitter Data

140 characters

Twitter lingo (e.g., #justinbieber, @msmith)

Internet acronyms (lol, omg) and emoticons (:D)

Misspellings

Slang/swearing
Approach: Supervised Machine Learning

Training Data (with labels)

Learn Model

Predict on Test Data (no labels)

@HVSVN

Don't fly @BritishAirways. Their customer service is horrendous.

Kyle Ayers

I've always said US Airways had the most leg room.

3:52 PM - 14 Apr 2014
123 RETWEETS 150 FAVORITES
Support Vector Machine (SVM)

Traditional approach: classifier (e.g., SVM) with manually selected features.
Features

N-grams (e.g., cleverness, wit, humor)

Part-of-speech (POS) tags

<table>
<thead>
<tr>
<th>Number</th>
<th>Tag</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.</td>
<td>CC</td>
<td>Coordinating conjunction</td>
</tr>
<tr>
<td>2.</td>
<td>CD</td>
<td>Cardinal number</td>
</tr>
<tr>
<td>3.</td>
<td>DT</td>
<td>Determiner</td>
</tr>
<tr>
<td>4.</td>
<td>EX</td>
<td>Existential <em>there</em></td>
</tr>
<tr>
<td>5.</td>
<td>FW</td>
<td>Foreign word</td>
</tr>
<tr>
<td>6.</td>
<td>IN</td>
<td>Preposition or subordinating conjunction</td>
</tr>
<tr>
<td>7.</td>
<td>JJ</td>
<td>Adjective</td>
</tr>
<tr>
<td>8.</td>
<td>JJR</td>
<td>Adjective, comparative</td>
</tr>
<tr>
<td>9.</td>
<td>JJS</td>
<td>Adjective, superlative</td>
</tr>
<tr>
<td>10.</td>
<td>LS</td>
<td>List item marker</td>
</tr>
<tr>
<td>11.</td>
<td>MD</td>
<td>Modal</td>
</tr>
<tr>
<td>12.</td>
<td>NN</td>
<td>Noun, singular or mass</td>
</tr>
<tr>
<td>13.</td>
<td>NNS</td>
<td>Noun, plural</td>
</tr>
<tr>
<td>14.</td>
<td>NNP</td>
<td>Proper noun, singular</td>
</tr>
<tr>
<td>15.</td>
<td>NNPS</td>
<td>Proper noun, plural</td>
</tr>
</tbody>
</table>
Neural Network Approach

“Recursive Deep Models for Semantic Compositionality over a Sentiment Treebank” (Socher et al, EMNLP 2013).
Stanford Sentiment Treebank

10,662 sentences from rottentomatoes.com
S
  /\  \
NP   VP
  /\  /\  \
 DT NN VP PP
 /\ /\ /\ /\ 
the woman with the telescope
  /\ /\ /\ /\ 
 VT NP saw the man
Recursive Neural Network

\[ p_2 = g(a, p_1) \]
\[ p_1 = g(b, c) \]

... not \quad very \quad good ...

a \quad b \quad c
Outline

1. App-building Tools
2. Current NLP Research
   a. Human Computation
   b. Sentiment analysis
   c. Machine translation
   d. Dialogue Systems
   e. Question answering
3. In-class Activity
Translation

**Source** Language (French)

**Target** Language (English)
I want to learn Spanish

Quiero aprender español
Challenges

Lexical ambiguity (e.g., “book the flight” vs. “read the book”)

Word order (e.g., SVO vs. SOV)

Syntactic ambiguity (e.g., “the woman saw the man with the telescope”)

Pronoun resolution (e.g., “Michael helped Mary. He was kind.”)

Tense (e.g., Spanish past tense includes both imperfect and preterite)
Traditional Approach

- Word- or phrase-based
- Syntax-based
- Semantics-based
Noisy Channel Model

1. Language model: $P(\text{target English sentence } e)$
2. Translation model: $P(\text{French source sentence } f \mid e)$

Best English sentence $e^* = \text{argmax } P(e \mid f)$

$= \text{argmax } P(f \mid e) P(e) / P(f)$

$= \text{argmax } P(f \mid e) P(e)$
Alignments

\[ e = \text{I like green eggs} \]

\[ f = \text{Me encanta huevos verdes} \]

\[ a = \{1, 2, 4, 3\} \]
IBM Translation Model

\[ P(f \mid e) = \sum P(f, a \mid e) \]

sum over all possible alignments \(a\)

\[ = \sum P(a \mid e) P(f \mid a, e) \]

chain rule

\[ = \sum P(a \mid e) \prod P(f_j \mid e_{aj}) \]

product over each word \(f_j\) and aligned \(e_{aj}\)

\[ = \sum c/(l + 1)^m \prod P(f_j \mid e_{aj}) \]

\(l + 1\) possible English words, \(m\) foreign words
Neural Machine Translation

Aims to build a single, large neural network for translation

Recurrent neural network (RNN) encoder-decoder
Learned Encodings

“Sequence to Sequence Learning with Neural Networks” (Sutskever et al, 2014)
Predicted Translations

Source:

An admitting privilege is the right of a doctor to admit a patient to a hospital or a medical centre to carry out a diagnosis or a procedure, based on his status as a health care worker at a hospital.

Translation:

Un privilège d’admission est le droit d’un médecin de reconnaître un patient à l’hôpital ou un centre médical d’un diagnostic ou de prendre un diagnostic en fonction de son état de santé.

Even better translation (with **attention** mechanism):

Un privilège d’admission est le droit d’un médecin d’admettre un patient à un hôpital ou un centre médical **pour effectuer un diagnostic ou une procédure**, selon son statut de travailleur des soins de santé à l’hôpital.

“Neural Machine Translation by Jointly Learning to Align and Translate” (Bahdanau et al, ICLR 2015)
Caption Generation

Results

- Man in black shirt is playing guitar.
- Construction worker in orange safety vest is working on road.
- Two young girls are playing with Lego toy.
- Boy is doing backflip on wakeboard.
Outline

1. App-building Tools
2. Current NLP Research
   a. Human Computation
   b. Sentiment analysis
   c. Machine translation
   d. **Dialogue Systems**
   e. Question answering
3. In-class Activity
Dialogue Systems
Semantic Decoder

Dialogue acts

“Book me a flight from New York to Boston.” -> book a flight

“No, I said Newark, not New York.” -> correction

Semantic tagging

“Book me a flight from **New York** to **Boston** on **Tuesday**.”
Dialogue Manager

Semantic Decoder
Observation

Dialogue State

Policy $\pi$

Possible Actions

Selected Action
Hey Siri, start playing The Truman Show

Okay Arron, where would you like to play it?

Apple TV
Living room

Arron’s iPad

“Okay glass”

Stop trying to strap me to your forehead, Danny. It won’t work.
Amazon Echo

Always ready, connected, and fast. Just ask.
Google Now
Outline

1. App-building Tools
2. Current NLP Research
   a. Human Computation
   b. Sentiment analysis
   c. Machine translation
   d. Dialogue Systems
   e. Question answering
3. In-class Activity
Question Answering

Quiz Bowl

Later in its existence, this polity’s leader was chosen by a group that included three bishops and six laymen, up from the seven who traditionally made the decision. Free imperial cities in this polity included Basel and Speyer. Dissolved in 1806, its key events included the Investiture Controversy and the Golden Bull of 1356. Led by Charles V, Frederick Barbarossa, and Otto I, for 10 points, name this polity, which ruled most of what is now Germany through the Middle Ages and rarely ruled its titular city.

Jeopardy

POP MUSIC: Their grandfather had a number 1 record in 1935; their father, Number 1’s in 1958 & 1961; and they hit number 1 in 1990.
(Answer: “Gunnar & Matthew Nelson”)

BEFORE & AFTER: The “Jerry Maguire” star who automatically maintains your vehicle’s speed.
(Answer: “Tom Cruise control”)
Dependency Tree Recursive Neural Network

“A Neural Network for Factoid Question Answering over Paragraphs” (Iyyer et al)
IBM Watson
DeepQA
Opportunity to work with Prof. Victor Zue!

Help build an online tool for learning spectrograms this spring

Email: zue@csail.mit.edu
Activity: Generate your own startup idea!

1. What is the problem?
2. What is your proposed solution?
   a. (as it relates to speech and language technologies)
3. Judging Criteria:
   a. Innovativeness, scalability, feasibility, market competition, business model
4. You are pitching to investors (us).
5. Describe your idea here (20 minutes):
   b. Team up (2-3 people)
   c. Limit to 3 slides
   d. Talk 3 minutes
Feedback on the class?

Your comments will help us improve the class!

Survey Link: https://goo.gl/hz4D0w

Thank you!