Natural Language Processing

Team 8

Eric Baisch 109936431    Noel Sicoco 109265689
John French 109653796    Shuqi Zhu 110654579
Jaykumar Patel 110255934

Dr. Anita Wasilewska    CSE 352 - Artificial Intelligence
Why we picked Natural Language Processing?

1. It’s used in several topics including Computer Science, Artificial Intelligence and Computational Linguistics (which uses statistical or rulebased modeling)

2. It’s used by several current applications.
   
   a. Google Search Engine
   
   b. Mobile Assistants
   
   c. Language Translations
Goal of Natural Language Processing

The goal of Natural Language Processing is to pass the Turing test. The Turing test states “A test for intelligence in a computer, requiring that a human being should be unable to distinguish the machine from another human being by using the replies to questions put to both.”

In order for a computer to pass the Turing test, it needs to convince more than 30% of the users that it is a human. To this date, there is around 3 programs to have passed the test with around 33% accuracy (The last test passing in 2014)
What is Natural Language Processing?

Natural Language Processing can be split into two parts:

a. Natural Language Understanding - Understanding spoken or typed language and acting on it based on meaning.

b. Natural Language Generation - Outputting results in a natural language format.

“The goal of the Natural Language Processing (NLP) [...] is to design and build software that will analyze, understand, and generate languages that humans use naturally, so that eventually you will be able to address your computer as though you were addressing another person” - Microsoft
Natural Language Understanding

1. Speech Recognition - Converting speech into a sentence.
   Match sounds to phenomes and match phenome sequences to words.

2. Syntactic Analysis - Determining the structure of the sentence.
   Nouns, Subjects, Verbs

3. Semantic Analysis - Determining the meaning of the sentence.
   "John loves Mary" → loves(john, mary)

4. Pragmatic Analysis - Determining the context/goal of the sentence.
Query: *Find me a flight from New York to California leaving today.*

### Query Data

<table>
<thead>
<tr>
<th>Airline</th>
<th>Time</th>
<th>Cost</th>
</tr>
</thead>
<tbody>
<tr>
<td>Delta</td>
<td>2:00PM</td>
<td>$400</td>
</tr>
<tr>
<td>Southwest</td>
<td>2:30PM</td>
<td>$500</td>
</tr>
</tbody>
</table>

### In Natural Language

“There are 2 flights available. The first flight is with Delta Airlines, leaves at 2pm, and costs $400. The second flight is with Southwest Airlines, leaves at 2:30pm, and costs $500.”
History of Natural Language Processing

1. Before the 1990s, NLP systems were hand-coded with symbolic and linguistic information.

2. After the 1990s, research on statistical NLP became more prominent.

3. Modern NLP uses both statistical and linguistic paradigms.
Parsey McParseface?

- “Parsey McParseface is built on powerful machine learning algorithms that learn to analyze the linguistic structure of language, and that can explain the functional role of each word in a given sentence.”

- Parsey was built using syntaxnet and neural-network Natural Language Processing framework for TensorFlow.
What does it do?

- It assigns grammatical labels to each word in a given sentence.
- Most accurate speech tagger in the world.
  - parser: 94% accuracy
  - Human: 96% accuracy

```
ROOT
  ▼
  | ▼
NSUBJ  DOBJ
  ▼  ▼
Alice  saw  Bob
  NOUN  VERB  NOUN
```
Facts about Parsey McParseface

Before Parsey McParseface:

   Long short-term memory (LSTM) networks dominate NLP.

What is it?:

   Globally normalized simple feed-forward networks without any recurrence.

Essential Algorithm:

1. Beam Search

2. Full backpropagation training of all neural network parameters based on the conditional random field (CRF) loss.
In greedy neural network parsing, the conditional probability distribution over decisions $d_j$ given context $(d_{1:j-1})$ is defined as follows:

The probability of a sequence of decisions $(d_{1:n})$ is to the left.
In contrast, a Conditional Random Field (CRF) defines a distribution $p_G(d_{1:n})$ as follows and $D_n$ is the set of all valid sequences of decisions of length $n$. $Z_G(\theta)$ is a global normalization term.

$$p_G(d_{1:n}) = \frac{\exp \sum_{j=1}^{n} \rho(d_{1:j-1}, d_{j}; \theta)}{Z_G(\theta)}, \quad (3)$$

where

$$Z_G(\theta) = \sum_{d'_{1:n} \in D_n} \exp \sum_{j=1}^{n} \rho(d'_{1:j-1}, d'_{j}; \theta)$$

Now the problem is to find:

$$\arg\max_{d_{1:n} \in D_n} p_G(d_{1:n}) = \arg\max_{d_{1:n} \in D_n} \sum_{j=1}^{n} \rho(d_{1:j-1}, d_{j}; \theta).$$

When beam sort again is used to find the approximation.
Training

- For each word and words around it, parser extract the set of features like prefix and the suffix.
- Put them into Data blocks and then concatenate them all together and send them to a feed forwards with lots of hidden layers which would then predict the probability distributions over set of possible POS tags.
- Going in order from left to right is useful because they can use previous word’s tag as a feature(attribute) in the next word.
**Example**

- “I saw her face and now I am a believer”
  - If we tag each word in the phrase individually without looking at the sentence as a whole we might tag saw as a certain verb.
  - If we look at this word in context of the sentence we quickly realize that it's a different verb.

- Google trained parsey to interpret sentences from left to right like a human being.
Parse Tree

- Head modifier construction to sort word dependencies this generates directed arc between words

- unprocessed part is called buffer, as the parser encounters words from left to right it pushes words on the stack.
Demo

Alice, who had been reading about SyntaxNet, saw Bob in the hallway yesterday.

NOUN   PRON   VERB   VERB   VERB   ADP   NOUN   .   VERB   NOUN   ADP   DET   NOUN   NOUN
Current Challenges in Natural Language Processing

- Languages other than English don’t have as much support yet, and the more different the grammar is from English, the more inaccurate NLP is with that language.

- There are many ways to word a sentence, and ideally an AI program should interpret the meaning correctly regardless of how a sentence is worded.

- Generally, only one sentence can be processed at a time, and connections between sentences cannot be made.

- Detection of sarcasm and other sentiments is still very difficult for AI.
Sources

The Essence of Artificial Intelligence by Alison Cawsey


https://en.wikipedia.org/wiki/Natural_language_processing

https://en.oxforddictionaries.com/definition/turing_test

https://research.googleblog.com/2016/05/announcing-syntaxnet-worlds-most.html


https://github.com/tensorflow/models/tree/master/syntaxnet