Whirlwind Tour of NLP Tasks

Winston Wu

9/10/2019
Last class

• Any interesting findings?
  • Funny sentences?
  • Cycles

• GPT-2
  • Trained on 40GB of web text
  • Try it out at https://gpt2.apps.allenai.org/
Today

• How do we evaluate a system’s understanding of language?
Turing Test

• Computing Machinery and Intelligence (Alan Turing, 1950)
  • Can machines think?
  • Imitation game
Text Classification

• Language detection

  Dutch: *En ik zeg jullie: wat zoeken jullie hier bij mij? Ik zoek jullie niet! Nee, ga nu weg!*
  Afrikaans: *En ek sê vir julle, wat soek julle hier by my? Ek soek julle nie! Nee, gaan nou weg!*

• Spam detection
• Sentiment Analysis

  the movie is shit
  the movie is the shit

• Fake news detection
Tokenization / Segmentation

• Sentence
  
  Mr. Smith, Jr. traveled to Mt. St. Helens on Sun. Apr. 2.

• Word

  thesetissues = “these tissues” or “the set issues”?

• Morpheme

  unambiguously = un + ambiguous + ly
Syntax: Sentence Structure

• Part of speech tagging

The/DET fires/NNS burned/VBZ the/DET building/NN

• Noun chunking

[The/DET fires/NNS] burned/VBZ [the/DET building/NN]
Tree Parsing

- Constituency parsing
- Dependency parsing
- Treebanks
Lexical Semantics: Word Meaning

• Word sense disambiguation

He fired at the man who was fired for starting the fire

• Relations
  • WordNet
Word Embeddings

• Word vectors / embeddings / distributional representation

“You shall know a word by the company it keeps” – J. R. Firth

• Old style: word2vec, GLoVE, fastText
• Contextual: ELMo, BERT, GPT, XLNet
Word Embeddings

- Multilingual Word Embeddings
  - Vector space alignment
  - Bilingual lexicon induction

- Sentence embeddings
- Document embeddings

MUSE: Multilingual Unsupervised and Supervised Embeddings
Information Extraction

• IE, Knowledge Base Population, Slot Filling
  • Semantic web

  Trump was born and raised in the New York City borough of Queens, and received his B.S. degree in economics from the Wharton School at the University of Pennsylvania.

• Named entity recognition
  blue above

• Entity linking
Semantics: Meaning

• Semantic role labeling / shallow semantic parsing

  The man/AGENT ate spaghetti/PATIENT with chopsticks/INSTRUMENT

• Meaning representation, event extraction

  The boy wants to go
  \( \exists e_1, e_2 \text{ want}(e_1, \text{boy}, e_2) \& \text{go}(e_2, \text{boy}) \)
Machine Translation

• 1949 – Warren Weaver “Translation” memorandum
  • Code breaking during WW2

• 1954 – Georgetown-IBM Experiment
  • Russian-English

• Meanwhile, growing interest due to computers becoming more powerful

• 2000s – Several breakthroughs
Machine Translation

• Phrase-based
• Syntax-based
• Neural
Automatic Speech Recognition

- Pipeline
- End-to-end
Text Generation

• Language modeling

• Text summarization
  • Extractive
  • Abstractive

• Text simplification

• Paraphrasing

• Grammatical error correction

• Dialog Systems / Conversational Agents
Question Answering

• 1968 – SHRDLU (Terry Winograd)
• Information retrieval

• Knowledge bases
• Open-domain QA
• Machine Comprehension
Language Understanding

• Textual Entailment / Natural Language Inference
  
  **Premise**: The dog is chasing the cat
  **Hypothesis**: The cat is sitting
  Entail, contradict, or neutral?

• Semantic Textual Similarity
• Common sense reasoning
Other

• Domain adaptation
• Multi-task learning / transfer learning
• Multi-modal
• Zero-shot, few-shot learning
Summary

• A lot of tasks!
• Groups
  • Label the entire text
  • Label parts of the text
  • Produce text
  • Show that you understand text
• http://nlpprogress.com/
Survey

1. What task(s) interested you the most? Briefly say why.
2. Would you be interested in a (lecture, pointer to reading, none) about human learning?
3. Questions, comments, concerns, suggestions