Computational Linguistics I

CMSC 723 / LING 723 / INST 725

MARINE CARPUAT
marine@cs.umd.edu
What is Computational Linguistics?

• Study of computer processing of natural languages

• Interdisciplinary field
  – Roots in linguistics and computer science (specifically, AI)
  – Influenced by many other fields
The field goes by various names...

• **Computational linguistics (CL)**
  – the science of doing what linguists do with language, but using computers.

• **Natural language processing (NLP)**
  – the engineering discipline of doing what people do with language, but using computers.

• Speech/language/text processing
• Human language technology/technologies
Science vs. Engineering

• What is the goal of this endeavor?
  – Understanding the phenomenon of human language
  – Building better applications

• Goals (usually) in tension
NLP State of the Art

Still a challenging problem!

But many useful applications already exist

AI’s Language Problem
“Machines that truly understand language would be incredibly useful. But we don’t know how to build them.”

MIT Technology Review
Will Knight, Aug 9, 2016
Today

• What does it mean for computers to process natural language?

• Why is this challenging?

• Class logistics
What’s a word?

• Break up by spaces, right?

  Ebay | Sells | Most | of | Skype | to | Private | Investors
  Swine | flu | isn’t | something | to | be | feared

• What about these?

  达赖喇嘛在高雄为灾民祈福
  ليبيا تحيي ذكرى وصول القذافي إلى السلطة
  百貨店、8月も不振 大手5社の売り上げ8～11%減
Morphological Analysis

• Morpheme = smallest linguistic unit that has meaning

• Morphemes are combined into words
  – duck + s = [N duck] + [plural s]
  – duck + s = [V duck] + [3rd person singular s]
  – happiness = [Adj happy] + [ness]
### Complex Morphology

In Turkish, from the root “uyu-” (sleep), the following can be derived…

<table>
<thead>
<tr>
<th>Turkish</th>
<th>English</th>
</tr>
</thead>
<tbody>
<tr>
<td>uyuyorum</td>
<td>I am sleeping</td>
</tr>
<tr>
<td>uyuyorsun</td>
<td>you are sleeping</td>
</tr>
<tr>
<td>uyuyor</td>
<td>he/she/it is sleeping</td>
</tr>
<tr>
<td>uyuyoruz</td>
<td>we are sleeping</td>
</tr>
<tr>
<td>uyuyorsunuz</td>
<td>you are sleeping</td>
</tr>
<tr>
<td>uyuyorlar</td>
<td>they are sleeping</td>
</tr>
<tr>
<td>uyuduk</td>
<td>we slept</td>
</tr>
<tr>
<td>uyudukça</td>
<td>as long as (somebody) sleeps</td>
</tr>
<tr>
<td>uyumalıyız</td>
<td>we must sleep</td>
</tr>
<tr>
<td>uyumadan</td>
<td>without sleeping</td>
</tr>
<tr>
<td>uyuman</td>
<td>your sleeping</td>
</tr>
<tr>
<td>uyurken</td>
<td>while (somebody) is sleeping</td>
</tr>
<tr>
<td>uyuyunca</td>
<td>when (somebody) sleeps</td>
</tr>
<tr>
<td>uyutmak</td>
<td>to cause somebody to sleep</td>
</tr>
<tr>
<td>uyutturmak</td>
<td>to cause (somebody) to cause (another) to sleep</td>
</tr>
<tr>
<td>uyutturtturmak</td>
<td>to cause (somebody) to cause (some other) to cause (yet another) to sleep</td>
</tr>
</tbody>
</table>

...
What’s a phrase?

• Coherent group of words that serve some function
  – Organized around a central “head”
  – The head specifies the type of phrase
• Examples:
  – Noun phrase (NP): the happy camper
  – Verb phrase (VP): shot the bird
  – Prepositional phrase (PP): on the deck
Syntactic Analysis

- Parsing: the process of assigning syntactic structure

```
[S [NP I] [VP saw [NP the man]]]
```

```
S
  /\NP
 /  \VP
  \ N
   \  V
    \ NP
     \  det
      \ N
       \  N
          \I
             saw
             the
             man
```
Semantic analysis

different words/structure, same meaning

– She needed to make a quick decision in that situation.
– The scenario required her to make a split-second judgment.

– I saw the man.
– The man was seen by me.
Semantic analysis

same words, different meaning

- I walked by the bank
  - ... to deposit my check.
  - ... to take a look at the river.

– Everyone on the island speaks two languages.
– Two languages are spoken by everyone on the island.
Discourse Analysis

• Discourse: how multiple sentences fit together

• Pronoun reference:
  – The dog wanted the bone, but Sam threw it away.

• Inference and other relations between sentences:
  – The bomb exploded in front of the hotel. The fountain was destroyed, but the lobby was largely intact.
Pragmatics and World Knowledge

- Interpretation of sentences requires context, world knowledge, speaker intention/goals, etc.
Why is CL/NLP hard?

easy...

Ambiguity!
Ambiguity at the word level

• Part of speech
  – [V Duck]!
  – [N Duck] is delicious for dinner.

• Word sense
  – I went to the bank to deposit my check.
  – I went to the bank to look out at the river.
Ambiguity at the syntactic level

- PP Attachment ambiguity
  - I saw the man on the hill with the telescope

- Structural ambiguity
  - I cooked her duck.
  - Visiting relatives can be annoying.
  - Time flies like an arrow.
Difficult cases...

• Requires world knowledge:
  – The city council denied the demonstrators the permit because they advocated violence
  – The city council denied the demonstrators the permit because they feared violence

• Requires context:
  – John hit the man. He had stolen his bicycle.
So how do humans cope?
How do computers cope?
I’d use the billion dollars to build a NASA-size program focusing on natural language processing (NLP), in all of its glory (semantics, pragmatics, etc).

Intellectually I think that NLP is fascinating, allowing us to focus on highly-structured inference problems, on issues that go to the core of “what is thought” but remain eminently practical, and on a technology that surely would make the world a better place.

http://www.reddit.com/r/MachineLearning/comments/2fxi6v/ ama_michael_i_jordan/
Social Impact

NLP experiments and applications can have a direct effect on individual users’ lives

Some issues
• Privacy
• Exclusion
• Overgeneralization
• Dual-use problems
Course Goals

By the end of the semester, you should be able to

• Design models and implement algorithms to address core NLP tasks
• Select linguistic representations that are appropriate for the problem you want to solve
• Read the NLP literature
Before next class...

• Read the syllabus
  http://www.cs.umd.edu/class/fall2016/cmsc723/

• Sign up for Piazza
  https://piazza.com/umd/fall2016/cmsc723/home

• Get started on HW1 (due by Wed Sep 7, 2pm)
  https://myelms.umd.edu

• Let me know
  • which religious holidays you will observe this semester
  • DSS letter of accommodation