SI485i : NLP

Missing Topics and the Future
Who cares about NLP?

• NLP has expanded quickly
  • Most top-tier universities now have NLP faculty (Stanford, Cornell, Berkeley, MIT, UPenn, CMU, Hopkins, etc)

• Commercial NLP hiring: Google, Microsoft, IBM, Amazon, LinkedIn, Yahoo

• Web startups in Silicon Valley are eating up NLP students

• Navy, DoD, NSA, NIH: all funding NLP research
What NLP topics did we miss?

- Speech Recognition
What NLP topics did we miss?

- Speech Recognition
What NLP topics did we miss?

- Machine Translation

On Tuesday Nov. 4, earthquakes rocked Japan once again

Des tremblements de terre ont à nouveau touché le Japon mardi 4 novembre
What NLP topics did we miss?

• Machine Translation

Start at ~6min in.
http://www.youtube.com/watch?feature=player_embedded&v=Nu-nlQqFCKg
What NLP topics did we miss?

- Machine Translation
- IBM Models (1 through 5)
Machine Translation

• How to model translations?

• **Words:** $P(casa \mid house)$
• **Spurious words:** $P(a \mid null)$
• **Fertility:** $P_n(1 \mid house)$
  • English word translates to one Spanish word
• **Distortion:** $P_d(5 \mid 2)$
  • The 2$^{nd}$ English word maps to the 5$^{th}$ Spanish word
Distortion

- Encourage translations to follow the diagonal...

- $P(4 \mid 4) \times P(5 \mid 5) \times \ldots$
Learning Translations

- Huge corpus of “aligned sentences”.
- Europarl
  - Corpus of European Parliament proceedings
  - The EU is mandated to translate into all 21 official languages
  - 21 languages, (semi-) aligned to each other

- \( P( \text{casa} | \text{house} ) = \) (count all casa/house pairs!)
- \( P_d(2 | 5) = \) (count all sentences where 2\(^{nd}\) word went to 5\(^{th}\) word)
Machine Translation Technology

- Hand-held devices for military
  - Speak english -> recognition -> translation -> generate Urdu

- Translate web documents

- Education technology?
  - Doesn't yet receive much of a focus
What NLP topics did we miss?

- Dialogue Systems

Do you think Anakin likes me?

I don’t care.
What NLP topics did we miss?

• Dialogue Systems

• Why? Heavy interest in human-robot communication.

• UAVs require teams of 5+ people for each operating machine
  • Goal: reduce the number of people
  • Give computer high-level dialogue commands, rather than low-level system commands
What NLP topics did we miss?

- Dialogue Systems

- Dialogue is a fascinating topic. Not only do we need to understand language, but now discourse cues:
  - Questions require replies
  - Imperatives/Commands
  - Acknowledgments: “ok”
  - Back-channels: “uh huh”, “mm hmm”

- Belief-Desire-Intention (BDI) Model
  - Beliefs: you maintain a set of facts about the world
  - Desires: things you want to become true in the world
  - Intentions: desires that you are taking action on
What NLP topics did we miss?

• Unsupervised Learning
What NLP topics did we miss?

• Unsupervised Learning

• Most of this semester used data that had human/gold labels.
  • Bootstrapping was our main counter-example: it is mostly unsupervised.

• Many many algorithms being researched to learn language and knowledge without humans, only using text.
El Fin

• Secret 1:
El Fin

- Secret 1:
  - I intentionally made our labs confusing
El Fin

- **Secret 1:**
  - I intentionally made our labs confusing

  *Under-defined tasks with unclear expected results*
El Fin

• Secret 1:
  • I intentionally made our labs confusing
    *Under-defined tasks with unclear expected results*

• Secret 2:
El Fin

• Secret 1:
  • I intentionally made our labs confusing
    \textit{Under-defined tasks with unclear expected results}

• Secret 2:
  • I tried to teach you skills that have nothing to do with NLP
El Fin

• Secret 1:
  • I intentionally made our labs confusing
    *Under-defined tasks with unclear expected results*

• Secret 2:
  • I tried to teach you skills that have nothing to do with NLP
    *Experimentation*
    *Error Analysis*
El Fin

• **Secret 1:**
  • I intentionally made our labs confusing
  *Under-defined tasks with unclear expected results*

• **Secret 2:**
  • I tried to teach you skills that have nothing to do with NLP
    *Experimentation*
    *Error Analysis*

• **Secret 3:**
El Fin

• Secret 1:
  • I intentionally made our labs confusing
  Under-defined tasks with unclear expected results

• Secret 2:
  • I tried to teach you skills that have nothing to do with NLP
    Experimentation
    Error Analysis

• Secret 3:
  • I appreciate the hard work you put into the class