### Structured Prediction for Language and Other Discrete Data 11-763

Introductory Lecture Taylor Berg-Kirkpatrick and Bhiksha Raj

Slides adapted from Chris Dyer, William Cohen, and Noah Smith

# **Course Information**

- Course website (link might not work yet): <u>http://structuredprediction.cs.cmu.edu</u>
- Piazza: should have gotten email!
- Canvas: check your email soon!

# **Course Requirements**

#### • Prerequisites:

- Upper division algorithms (dynamic programming)
- Strong background in machine learning (e.g. 701)
- Programming skills
- Work and Grading:
  - 5 Assignments (mostly coding, some written)

# The Book

MORGAN & CLAYPOOL PUBLISHERS

#### Linguistic Structure Prediction

Noah A. Smith

Synthesis Lectures on Human Language Technologies

Graeme Hirst, Series Editor

- Linguistic Structure Prediction
- Available in electronic form (free at CMU) and print form.

### **Other Announcements**

#### • Course Contacts:

- Website: slides, readings, assignments
- Piazza: announcements, discussion forum
- Canvas: turning in assignments
- Enrollment:
  - We'll try to take everyone who meets the requirements
- Late day policy:
  - 5 late day quota to use whenever you like
  - Subtract 25% credit for every additional day after quota

# A Little Bit of History

1935: Zipf's law

1940s & 1950s: empiricism: Shannon, Weaver, Harris, Yngve

# George Kingsley Zipf, 1935



 $p(w) \approx$  $\frac{1}{rank(w)}$ 

 Heavy tail in word distributions

# Claude Shannon, 1948



- Father of information theory
- Entropy: a mathematical measure of uncertainty
- Information can be encoded digitally; questions include how to encode information efficiently and reliably.
- Huge impact on speech recognition (and space exploration and digital media invention and ...)

### Warren Weaver, 1949

• "One naturally wonders if the problem of translation could conceivably be treated as a problem in cryptography. When I look at an article in Russian, I say: 'This is really written in English, but it has been coded in some strange symbols. I will now proceed to decode."



# Zellig Harris, 1940s and forward



- Centrality of data for linguistic analysis
- Transformations (a step toward computational models of language)
- Heavy use of mathematics in linguistics

# Victor Yngve, 1958

- Early computational linguist
- Showed "depth limit" of human sentence processing - restricted left branching (but not right)
- Early programming language, COMIT, for linguists (influenced SNOBOL)
- Random sentence generation (in the 1950s)

# A Little Bit of History

1935: Zipf's law

1940s & 1950s: empiricism: Shannon, Weaver, Harris, Yngve

1960-1985: rationalism/representations/formalisms/syntax/unapplied AI

- 1962: ACL (then MTACL) begins
- 1964-6: ALPAC report, MT winter, Bar-Hillel leaves the field

1980: ICML begins

- ~1985: statistical and information theoretic methods catch hold again in NLP, in part due to their success in ASR
  - This has continued unabated for 25+ years, with help from Moore's Law-type phenomena
- 1986: LTI founded (then called "CMT")
- 1993: "Very Large Corpora" workshops start at ACL
- 1996: EMNLP conference starts

~1997: Lafferty and Rosenfeld start teaching "Language and Statistics" at CMU

1998-early 2000s: Internet boom, commercial language technologies becoming viable

- ~2003: MLD founded (then called "CALD")
- 2004: Cohen starts teaching "Information Extraction"
- 2006: Smith starts teaching "Language and Statistics 2"
- 2011: Cohen and Smith start teaching "Structured Prediction"

# What is Structured Prediction?

Having observed some information (input) ...

- Binary classification: predict a coin toss (given some information)
- Multi-class: predict which side of a die (given some information)
- Structured prediction: choose among a very large number of complex outcomes.
  - Large means "exponential in the size of the input."

# E.g., (Part of Speech) Tagging



#### E.g., Segmentation into Words

第二阶段的奥运会体育比赛门票与残奥会开 闭幕式门票的预订工作已经结束,现在进入 门票分配阶段。在此期间,我们不再接受新的 门票预订申请。

#### E.g., Segmentation within Words

uygarlaştıramadıklarımızdanmışsınızcasına

"(behaving) as if you are among those whom we could not civilize"

# E.g., Segmentation and Tagging





#### E.g., Predicate-Argument Structures



stranding thing

# E.g., Alignments

Mr President, Noah's ark was filled not with



production factors , but with living creatures .

Produktionsfaktoren, sondern Geschöpfe.

## Gene Finding and Analysis



Slide due to E. Xing

# Phylogenetic Relationships



#### **Image Segmentation**







from Nowozin and Lampert (2010)

# Implications of "Going Structured"

- All aspects of training and testing become more complex:
  - Designing a model
  - Prediction algorithms (once you have a model)
  - Learning your model from data
  - Measuring "error" of a prediction
- Machine learning helps with "mental hygiene"!
  - Principles that will help you explain and understand your methods
  - Generic optimization algorithms
  - Formal guarantees (sometimes)
  - Baselines when you're tackling a new problem

# The Structured Prediction Way

- 1. Formally define the inputs and outputs.
- 2. Identify a scoring function over input-output pairs, and an algorithm that can find the maximum-scoring output given an input.
- 3. Determine what data can be used to learn to predict outputs from inputs, and apply a learning algorithm to tune the parameters of the scoring function.
- 4. Evaluate the model on an objective criterion measured on unseen test data.

# Topics

- Inference (ch. 2, 5)
- Learning from Complete Data (ch. 3)
- Learning from Incomplete Data (ch. 4)

## SPFLODD and Other Classes



# Homework for Thursday

• Read *LSP*, preface and chapter 1.