Wrap Up

Mausam

Grading

- 23% project
  - 2 proposal
  - 3 mid-project milestone
  - 6 peer grading
  - 6 Mausam grading
  - 4 TA grading
  - 2 grading
- 30% final exam
- 20% midterm
- 9% A3
- 9% A2
- 9% A1

Project Presentation

Evaluation Guidelines

- Clarity of presentation
- Significance of contribution
- Correctness of approach
- Strength of results
- Impressiveness of demo

- Mean & Median
  - Between 5 and 7
- Std dev >=1

Discussion
Chomsky vs. Norvig

Science <-> Engineering

- Goal of science: understanding laws of nature
  - create theories …
  - insight
- Goal of engineering: approximating unanalyzed data
  - create tools
- What is success for cognitive science? For engineering?

Is Grammar Deterministic?

- Grammar evolves
- People use non-grammatical constructions
  - Quake – intransitive/transitive?
- Grammaticality is region-specific
- All grammars leak!

Accurate Description <-> Insight

- Darwin
  - False facts are highly injurious to the progress of science, for they often endure long; but false views, if supported by some evidence, do little harm, for every one takes a salutary pleasure in proving their falseness.
- Feynman
  - Physics can progress without the proofs, but we can't go on without the facts.
- Rutherford
  - All science is either physics or stamp collecting.
- Chomsky
  - You can also collect butterflies and make many observations. If you like butterflies, that's fine; but such work must not be confounded with research, which is concerned to discover explanatory principles.

Accurate Description <-> Insight

- Millikan
  - Science walks forward on two feet, namely theory and experiment ... Sometimes it is one foot that is put forward first, sometimes the other, but continuous progress is only made by the use of both.

Data modeling <-> Algorithmic modeling

- nature can be described as a black box that has a relatively simple underlying model which maps from input variables to output variables (with perhaps some random noise thrown in).

- nature's black box cannot necessarily be described by a simple model. Complex algorithmic approaches are used to estimate the function that maps from input to output variables, but we have no expectation that the *form* of the function that emerges from this complex algorithm reflects the true underlying nature.

Chomsky's objection

- algorithmic modeling produces a form that, while accurately modeling reality, is not easily interpretable by humans, and makes no claim to correspond to the generative process used by nature. In other words, algorithmic modeling describes what *does* happen, but it doesn't answer the question of *why*.

Generation <-> Interpretation

- Chomsky vs. Shannon
  - Generation does not expose the inherent uncertainty
  - Interpretation cannot be done without taking it into account

Key Points of the Course

Key Points: NLP

- Challenges
  - Ambiguity, Ambiguity, Ambiguity, Sparsity
- Words: Morphology
- Sentences: Syntax
  - POS tagging, NP chunking, Parsing
- Sentences/Documents: Semantics
  - Words/Bigrams encode meanings, but are also sparse
  - Distributional Semantics, Shallow semantics
  - Patterns: bootstrapping
- Documents: Coreference, Discourse
- Applications
  - Information Extraction, Machine Translation, Summarization, Dialog

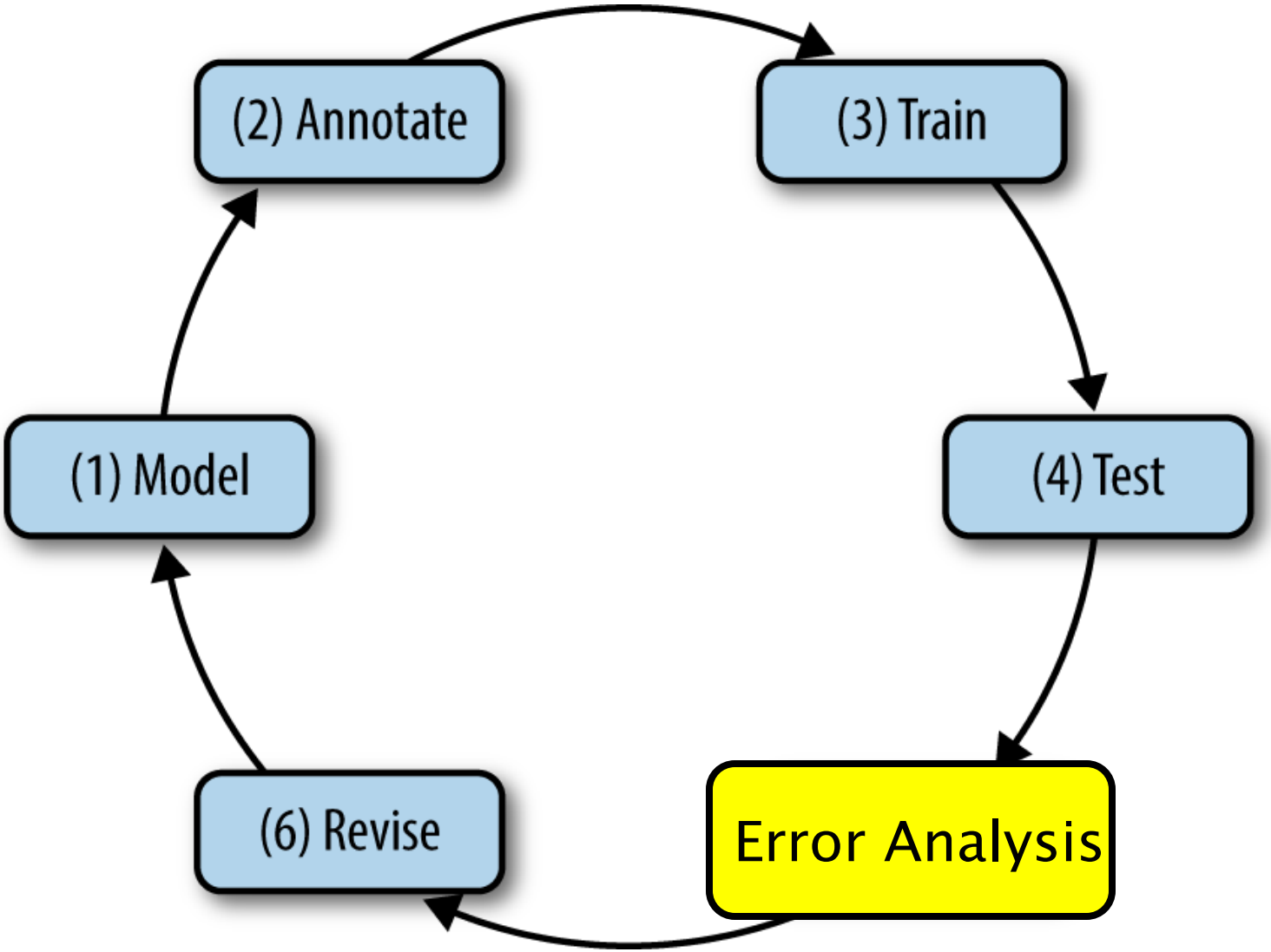
Key Points: Modeling

- Models
  - Logical vs. Probabilistic vs. Neural
  - Representation: Bag of Word-level, Linear, Tree
  - Generative vs. Discriminative
  - Smoothing / Regularization
  - Pipeline vs. Joint inference
- Training Data
  - A lot: Supervised
  - None: Unsupervised
  - A little: Semi-supervised
  - Distantly supervised
  - Active
- Features, Features, Features: local or global
- Learning Representations

Key Points: Data Insights

- Data annotation: be a linguist!
- Crowdsourced Data Curation
- Data harvesting
  - Example: bootstrapping
  - Example: Summarization
- Indirectly related data
  - Example: distant supervision

Key Points: ML Cycle



Neural Models

- Shallow NNs
  - Bag(words)
- Convolutional NNs
  - Handle bag (fixed length n-grams)
- Recurrent NNs
  - Handle small variable length histories
- LSTMs/GRUs
  - Handle larger variable length histories
- Bi-LSTMs
  - Handle larger variable length histories and futures
- Recursive NNs
  - Handle variable length partially ordered histories

Neural Models (contd)

- Hierarchical Recurrent NNs
  - RNN over RNNs (e.g., HRED)

- Neural language models
- Conditioned language models
  - Encoder-Decoder Models

- Attention models
  - attach non-uniform importance to histories based on evidence (question)
  - maybe even more important than recurrence?

Probabilistic or Neural?

- Amount of supervised data
- Ability to use background knowledge
- Need/effectiveness of specific constraints
- Need/effectiveness of human-designed features
- Interpretability
- Joint inference
- Ease of training

- Probabilistic + Neural