Natural Language Processing

* Co-reference Resolution
* First identity words’ type and then their relation to one another
  + Semantics passing ->element of data
* Unstructured text -> database (tables)
* Structure of Knowledge -> Archives of document
  + Stock prices
* Sentiment/Opinion Analysis
* Product design

Brainstorm

* NLP Applications
  + Music (sequences of notes)
  + Article Generator
  + Translation
* Different interpretation -> ambiguity in text
  + PP attachment ambiguity
  + Pronoun reference ambiguity
  + Changing language – cyber elingo
* NLP Models – Brittle
* Word embeddings
  + v(man) – v(woman) + v(uncle) ~ v(aunt)
    - Direction of relations
* Human -> machine learning exemplifies biases
* Gender Bias in Coref
  + Model -> constrain
* Machine learning – label space -> human annotation
* Prerequisites
  + Linear algebra/calculus/probability
  + CS146 – intro to machine learning
    - Natural language processing

kwchant.net

Teach human languages

* Expression: understand the entities
  + Grouping of words -> decisions – complex
    - Word phases (divide each by its part of speech)
  + Understand the sent -> convert to logic form
* Information Extract: issues/ complications
  + Unstructured text to database entries -> map to chart
* Processing of structured knowledge
* Uses: machine translation, sentiment/ opinion analysis
* Challenges (model training – difficult to interface)
  + Languages – compositional (structure is depended of one another)
  + Ambiguity
    - Lexical ambiguity: words w/ multiple meanings
      * “I saw her duck”
    - Syntactic ambiguity
      * Propositional phrase (PP) attachment ambiguity
      * Others
        + Conjunction, noun group structure
      * “San Jose cops kill man with knife”
        + Attachment of the “knife” to different subjects
    - Semantic ambiguity
      * After resolving syntax and meanings -> diff ways of reading
    - Anaphoric ambiguity
      * Referring something previous mentioned
      * Pronoun reference
  + Language changes (NOT static) -> increased in database
* Problems
  + Word embedding -> common trends/ data of our society -> biased
    - EX: sexist (male and female positions)
      * Google W2V embedding trained from news
  + Human bias in structured prediction models – according to the trainings
  + Dataset gender bias (ex. 33% male to 66% female for cooking)
    - Results: exemplified
      * 16% male and 84% female
      * Gender bias in Coref -> make the model coherence to the dataset
* Machine learning: input -> through statistical model -> label space
  + Human annotation
* Prereq: linear algebra + calc + prob
  + CS146 – Intro to ML
  + NLP intros