Processing Domain-Specific Language with Hybrid Technologies

Jacob Collard – jacob.collard@nist.gov

April 15, 2021

NIST Information Technology Lab

Domain-Specific Concepts

A Technical Abstract (Davis et al 2019)

A standard approach to evaluating language models analyzes how models assign probabilities to valid versus invalid syntactic constructions (i.e. is a grammatical sentence more probable than an ungrammatical sentence). Our work uses ambiguous relative clause attachment to extend such evaluations to cases of multiple simultaneous valid interpretations.

Domain-Specific Concepts

A Technical Abstract (Davis et al 2019)

A standard approach to evaluating language models analyzes how models assign probabilities to valid versus invalid syntactic constructions (i.e. is a grammatical sentence more probable than an ungrammatical sentence). Our work uses ambiguous relative clause attachment to extend such evaluations to cases of multiple simultaneous valid interpretations.

Goals and Challenges

Goals

- Identify concepts from domain-specific texts
- Identify relations between concepts
- Apply concepts and their relations to practical problems

Goals and Challenges

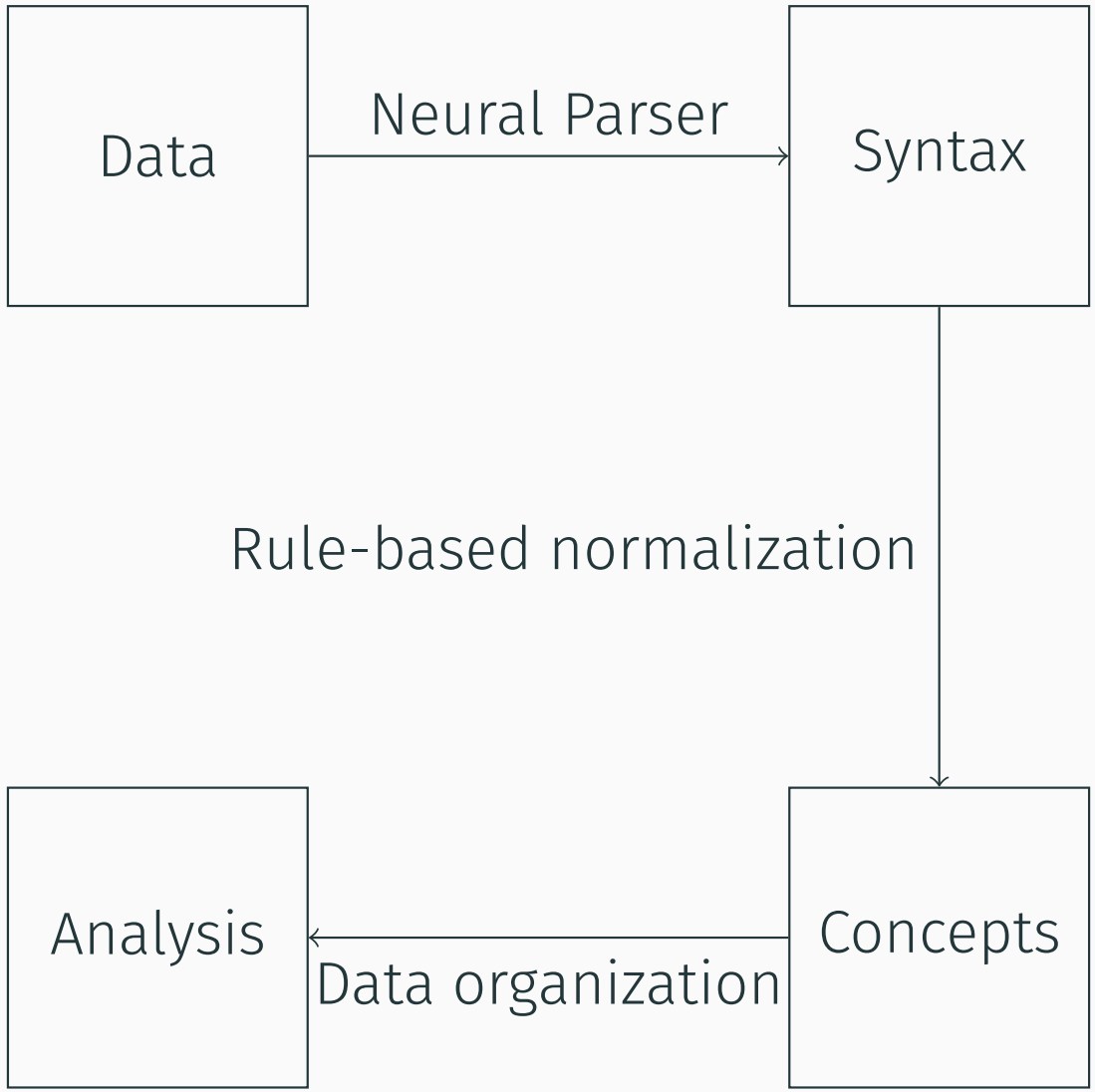
Goals

- Identify concepts from domain-specific texts
- Identify relations between concepts
- Apply concepts and their relations to practical problems

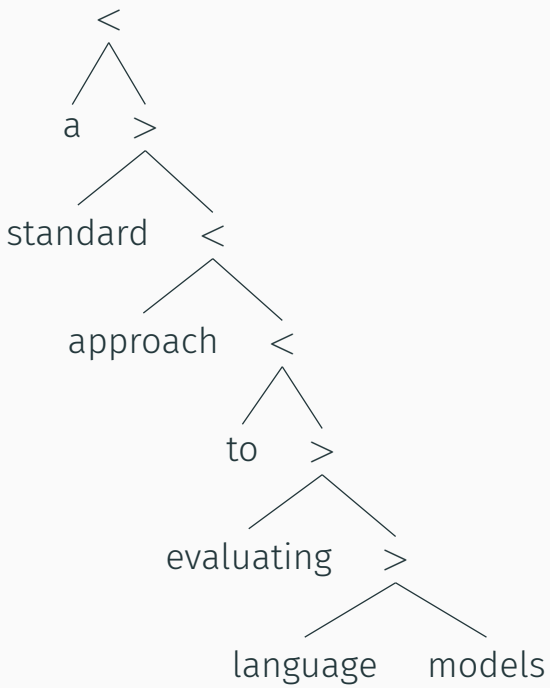
Challenges

- Symbolic processing is often not robust
- Neural networks are hard to configure

General Approach



Parsing and Normalization

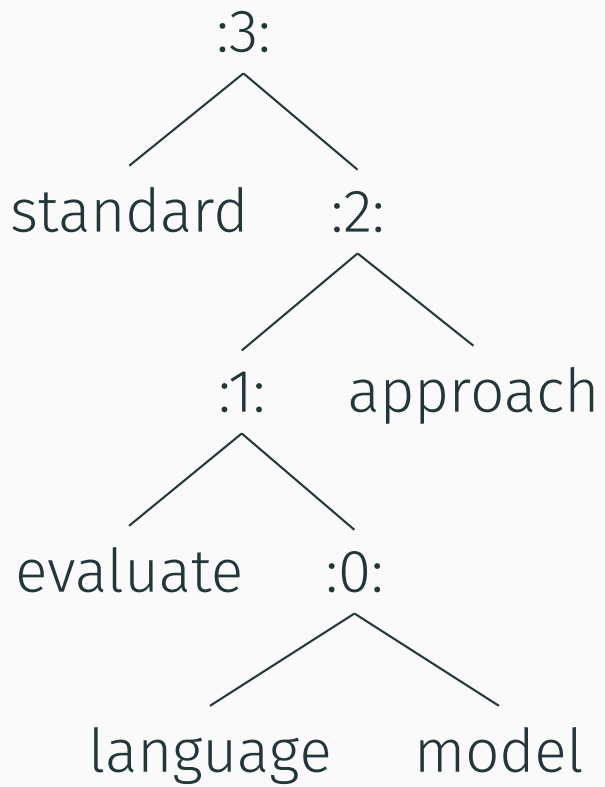


A standard approach to evaluating language models

Standard approaches to language model evaluation

Standard language model evaluation approaches

~~Approaches to evaluating standard language models~~



standard:3:evaluate:1:language:0:model:2:approach

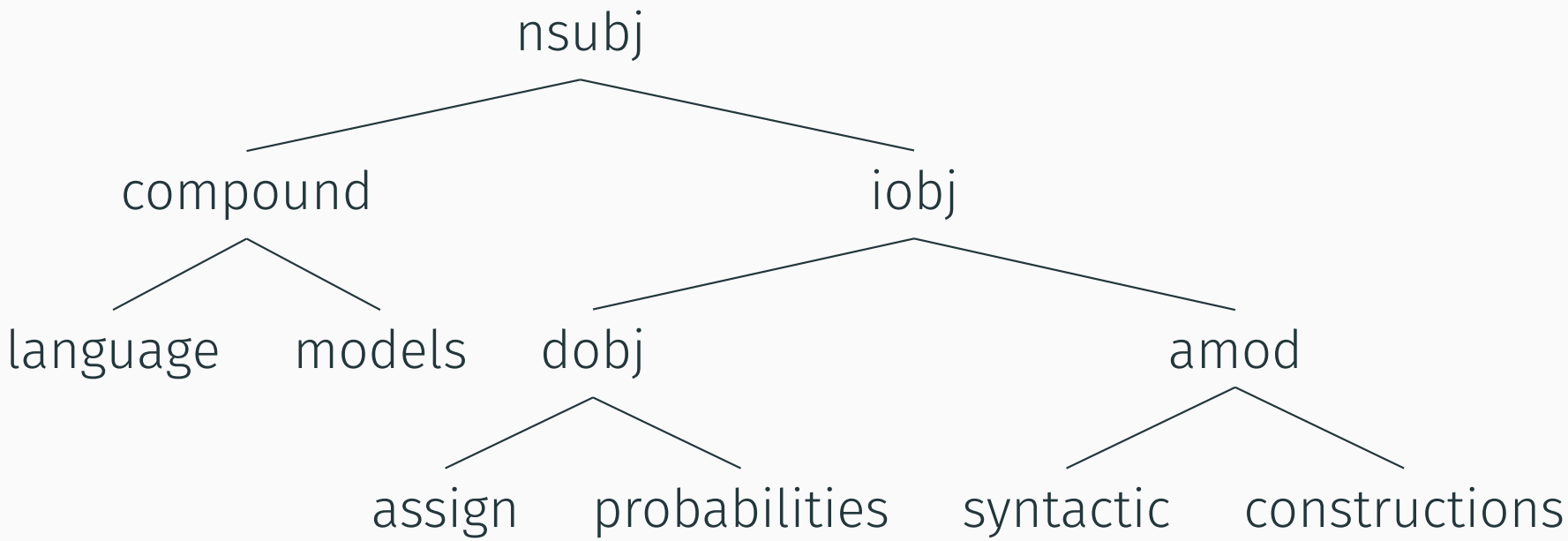
Semantic Structure

standard :3: evaluate :1: language :0: model :2: approach

Taxonomy

- language model
  - neural language model
  - masked language model
  - causal language model
  - large language model
- sentence
  - grammatical sentence
  - ungrammatical sentence
  - interrogative sentence

Identifying Relations



(language models, assign probabilities, syntactic constructions)

Future Work

- Concepts identified in this way can be combined with topic models and language models
- User interfaces for exploring concepts and relations are critical
- Evaluation is also critical – and difficult
- Users and developers can curate results
- Combining neural and symbolic models provides both robustness and power