

Natural Language Processing

Sargur N. Srihari

srihari@cedar.buffalo.edu

This is part of lecture slides on [Deep Learning](http://www.cedar.buffalo.edu/~srihari/CSE676):
<http://www.cedar.buffalo.edu/~srihari/CSE676>

Topics in NLP

- Overview

1. N-gram Models
2. Neural Language Models
3. High-dimensional Outputs
4. Combining Neural LMs with n-grams
5. Neural Machine Translation
6. Attention Models
7. Historical Perspective

What is NLP?

- Computers process *unambiguous languages*
 - E.g., C++, Python allow unambiguous parsing of programs
- Naturally occurring human languages, e.g, English are ambiguous
 - They defy formal categorical description
- NLP includes applications: Machine Translation
 - Learner reads sentence in a language and emits a sequence in another
- NLP is based on language models
 - That define a probability distribution over sequences of words, characters or bytes of natural language

Why to use probability in NLP?

- Natural language is ambiguous, because
 - Its use is situated in a world context
 - We say little that is needed to be understood
- Tasks involve reasoning about meaning
 - Using world knowledge not in source text
 - We need quantitative techniques
 - Which move away from unrealistic categorical assumptions of earlier formal linguistic theory
- Models that assign probabilities to sequences of words are called language models or LMs

Probability in machine translation

- Chinese source sentence
 - 他向记者介绍了主要内容
 - He to reporters introduced main content
- Potential rough translations
 - he introduced reporters to the main contents of the statement
 - he briefed to reporters the main contents of the statement
 - **he briefed reporters on the main contents of the statement**
- Last one is most probable phrase

Sequential models for natural language

- Natural language is sequential
 - Often regarded as a sequence of words, rather than a sequence of characters or bytes
- High dimensionality
 - Total no of words is very large
 - So, word-based language models must operate on an extremely high-dimensional and sparse discrete space
 - Several strategies have been developed to make models of such a space efficient
 - Both computationally and statistically

Language model as distribution

- An LM defines a probability distribution over tokens in natural language
 - A token is a word, character or byte
- Earliest successful LMs were based on fixed-length sequences of tokens called n -grams
 - Simplest n -gram models assign probabilities to sentences and sequences of words

Neural Network Language Models

- Generic Methods
 - Very generic neural network techniques can be successfully applied to NLP
- Domain-specific Methods
 - To achieve high performance and scale well to large applications, domain-specific strategies become important
 - To build an efficient model, we must use techniques specialized to sequential data