Natural Language Processing

Mairi Hallman

June 2024

Contents

1   Introduction

Natural Language Processing (NLP) methods may either take in unstructured natural language as input, or produce it as output.

Much like the humans who use them, natural languages are as abiguous as they are variable. Consider the following examples:

I love cooking, my family, and my cat. I love cooking my family and my cat.

The addition of two commas completely changes the meaning of the sentence.

Humans excel at using language (expression, perception, and nuance), but fall short when it comes to understanding and explaining the axioms that govern language. Without these axioms, it is very difficult to utilize algorithms that require a more formal lingustic framework. As a consequence, computers typically have a hard time interpreting and outputting natural language.

Machine learning methods, particularily subervised learning algorithms, can be quite effective when defining "good" rules is a challenge. However, annotating outputs is simple compared to what humans can do (even if it is resource-intensive).

Challenge: input is variable and/or ambiguous
rules are unknown and/or poorly defined
natural languages are • **discrete:**

2   Learning Basics and Linear Models

3   Working with Natural Language Data

4   Case Studies of NLP Features

5   From Textual Features to Inputs

6   From Textual Features to Inputs

7   Language Modelling

8   Pre-Trained Word Embeddings

9   Using Word Embeddings

10   Case Study of Sentence Meaning Inference