

Natural Language Processing

An Introduction

Shannon Hateley
UC Berkeley CDIPS 2016
Thursday, July 07, 2016

Outline

Natural Language Processing

- General Idea
- Common Techniques
- Tutorial

NLP: artificial intelligence meets linguistics

Using AI to analyze, understand, and generate human languages so that computers can interface with written and spoken word.

"Baby swallows fly."

Whether "swallows" or "fly" is the verb also determines whether "baby" is used as a noun or an adjective.

Context plays a huge part in human language. This must be accounted for in order for computers to correctly interpret natural language.

Some NLP applications/ problems

- Sentence boundary detection ('Dr.', 'e.g.')
- Tokenization – identifying individual elements ('10 mg/day', 'New York-based')
- Part-of-speech assignment (e.g. gerunds 'swimming is fun')
- Sarcasm and humor can be difficult
- Automatic summarization
- Sentiment analysis

Common Steps

1. Get data
2. Clean data and preprocess text
3. Create features – identify words, etc.
4. Apply some type of Machine Learning

tutorial

Kaggle Bag of Words Meets Bags of Popcorn

<https://www.kaggle.com/c/word2vec-nlp-tutorial>

Get the interactive ipython notebook at my
github: <https://github.com/hateley>

NLP_tutorial

Amazon EC2

Tutorial:

- <http://angus.readthedocs.io/en/2015/amazon/index.html>

In your EC2 instance, you'll pretty much always type:

```
apt-get update
```

```
apt-get -y install (whichever programs you want)
```

Examples from genomics can be seen here:

```
http://angus.readthedocs.io/en/2015/index.html
```