

Language Technology Group

**Maja Buljan, Vinit Ravishankar, Ildikó Pilán, Jeremy Barnes,
Samia Touileb, Lilja Øvrelid, Erik Velldal, Stephan Oepen, Jan
Tore Lønning**

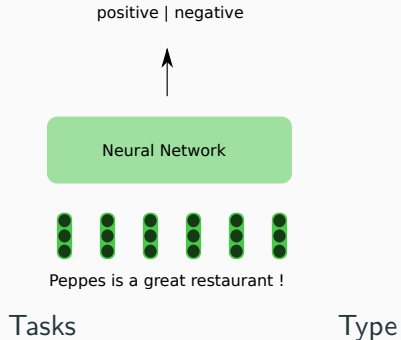


2-minute overview of NLP

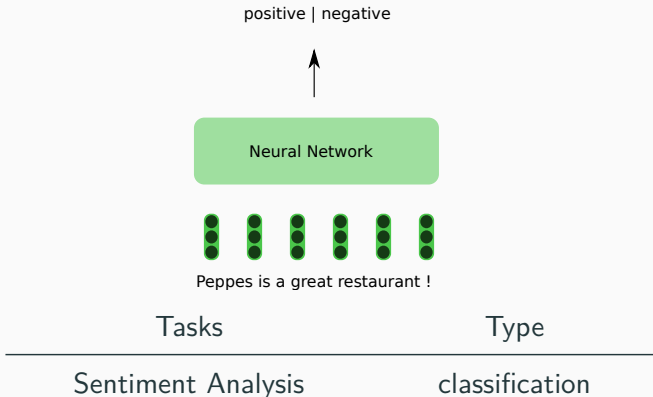
Tasks

Type

Tasks



Tasks



Tasks

B O O O O O O



Neural Network



Peppes is a great restaurant !

Tasks

Type

Sentiment Analysis

classification

Tasks

B O O O O O O



Neural Network



Peppes is a great restaurant !

Tasks

Type

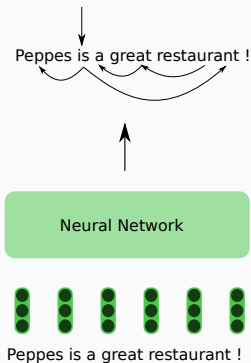
Sentiment Analysis

classification

Named Entity Recognition

sequence labeling

Tasks



Tasks

Type

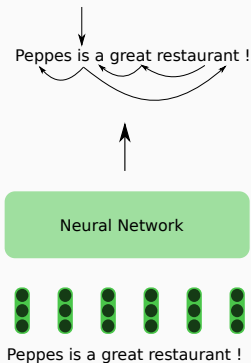
Sentiment Analysis

classification

Named Entity Recognition

sequence labeling

Tasks



| Tasks | Type |
|--------------------------|-----------------------|
| Sentiment Analysis | classification |
| Named Entity Recognition | sequence labeling |
| Parsing | structured prediction |

Tasks

Peppes jatetxe bikaina da!



Neural Network



Peppes is a great restaurant !

Tasks

Type

Sentiment Analysis

classification

Named Entity Recognition

sequence labeling

Parsing

structured prediction

Tasks

Peppes jatetxe bikaina da!



Neural Network



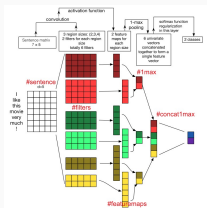
Peppes is a great restaurant !

Tasks

Type

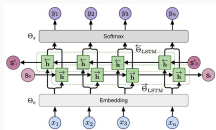
| | |
|--------------------------|-----------------------|
| Sentiment Analysis | classification |
| Named Entity Recognition | sequence labeling |
| Parsing | structured prediction |
| Machine Translation | structured prediction |

Models



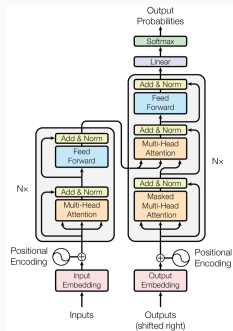
(a)

CNNs



(a)

RNNs



(a)

Transformers

- Unlabeled (English) - Terrabytes
- Labeled (English) - 1000 - 5M depending on task

Exciting Areas

Exciting Areas

- NLP has become good enough for deployment*

Exciting Areas

- NLP has become good enough for deployment*
- Better utilization of unlabeled data

Exciting Areas

- NLP has become good enough for deployment*
- Better utilization of unlabeled data
- Cross-lingual approaches

Exciting Areas

- NLP has become good enough for deployment*
- Better utilization of unlabeled data
- Cross-lingual approaches
- More interest in low-resource languages

Exciting Areas

- NLP has become good enough for deployment*
- Better utilization of unlabeled data
- Cross-lingual approaches
- More interest in low-resource languages

Challenges

Challenges

1. Lack of data for languages other than English

Challenges

1. Lack of data for languages other than English
2. Zipfian distribution of language phenomena

Challenges

1. Lack of data for languages other than English
2. Zipfian distribution of language phenomena
3. Lack of interpretability

Challenges

1. Lack of data for languages other than English
2. Zipfian distribution of language phenomena
3. Lack of interpretability
4. How to incorporate world knowledge

Challenges

1. Lack of data for languages other than English
2. Zipfian distribution of language phenomena
3. Lack of interpretability
4. How to incorporate world knowledge
5. Much of our data is very noisy

Challenges

1. Lack of data for languages other than English
2. Zipfian distribution of language phenomena
3. Lack of interpretability
4. How to incorporate world knowledge
5. Much of our data is very noisy
6. Oversized models