Named Entity Recognition

on datasets with little annotated data

Pius von Däniken

SpinningBytes AG

Nicole Falkner

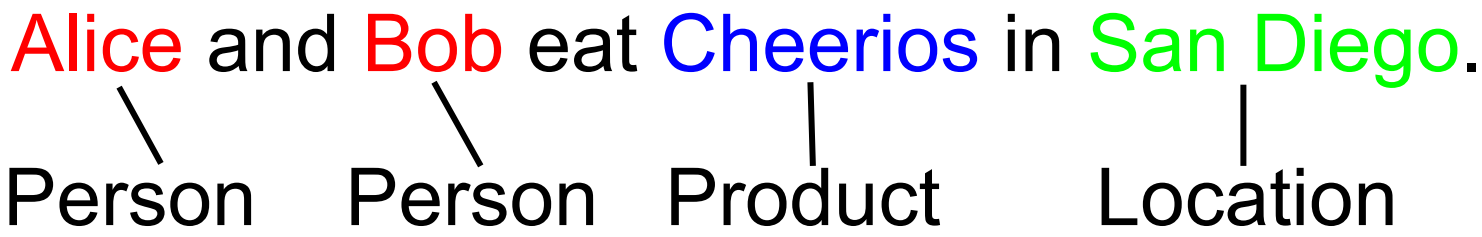
ZHAW

Stefano Dolce

ZHAW

SPINNINGBYTES

Example


Alice and Bob eat Cheerios in San Diego.
Person
Person
Product
Location

Example

Motivation

- Research
- Preprocessing Step for downstream tasks
  - Summarisation, Information Extraction
  - Chatbots!

Datasets

CoNLL 2003

Dataset of the shared task at the **C**onference **o**n **N**atural **L**anguage **L**earning

~15’000 annotated sentences from the Reuters corpus rcv1

location	miscellaneous
organisation	person

Sample	
Baker	I-person
made	O
secret	O
trip	O
to	O
Syria	I-location
in	O
March	O
1995	O
.	O

WNUT 2016

Dataset of the shared task at the **W**orkshop on **N**oisy **U**ser-generated **T**ext

~2400 annotated tweets

facility	geo-location
movie	music artist
organisation	other
person	product
sports team	tv-show

Sample	
Rob	B-person
Halford	I-person
at	O
Judas	B-musicartist
Priest	I-musicartist
Press	O
Conference	O
In	O
Hollywood	B-geoloc
<URL>	O

CAp 2017

Dataset of the shared task at the “Conférence sur l’Apprentissage Automatique”

~1900 annotated tweets

event	transportline
media	
facility	geo-location
movie	music artist
organisation	other
person	product
sports team	tv-show

Sample	
Polémique	O
:	O
Nicolas	B-person
Sarkozy	I-person
et	O
“	O
les	O
Gaulois	B-other
“	O
<URL>	O

State of the Art

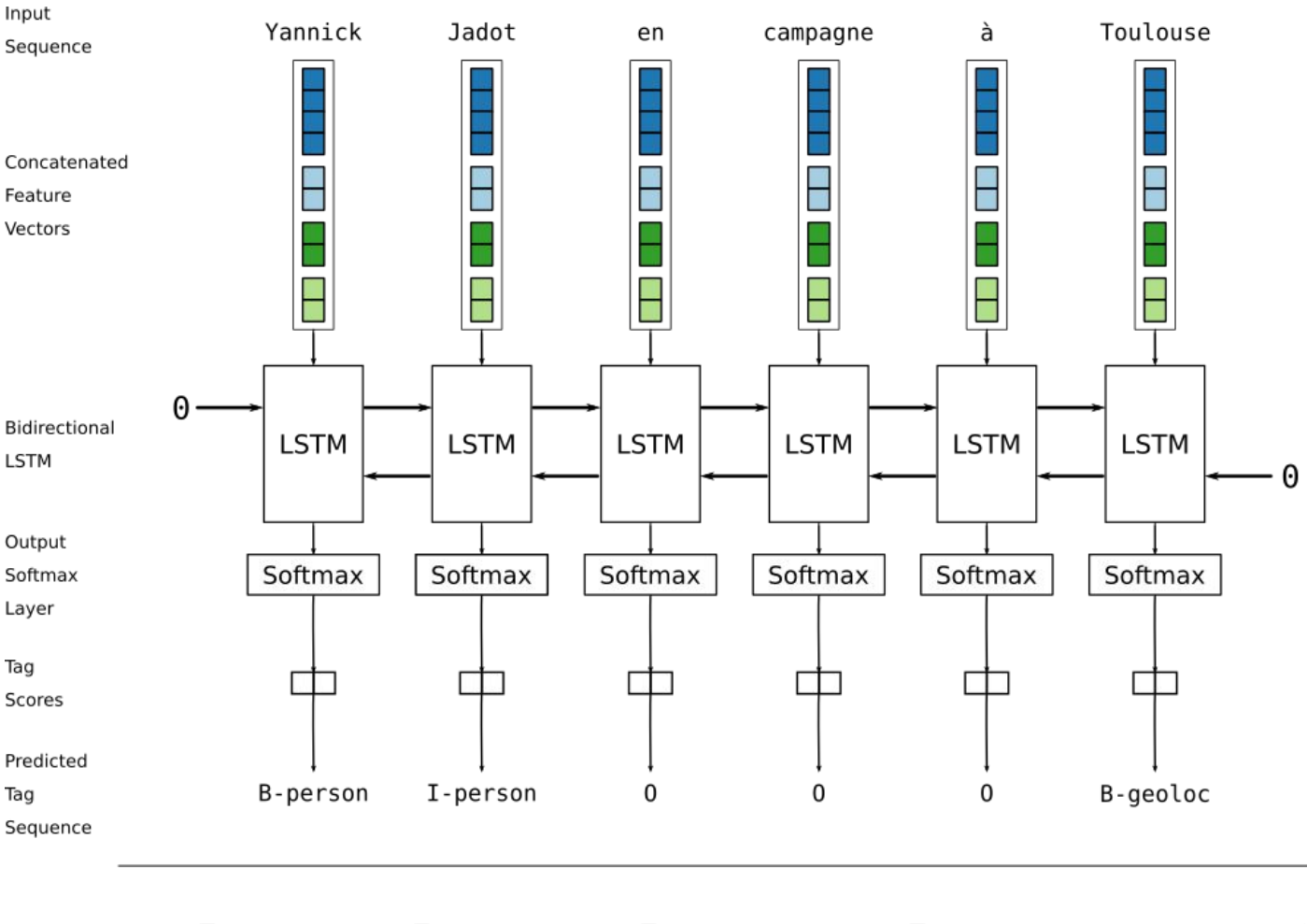
Dataset	Language	Domain	F1 score	Authors
CoNLL-2003	English	News	91.62	Chiu & Nichols, 2015
WNUT-2016	English	Twitter	52.41	Limsopatham & Collier, 2016
CApNER-2017	French	Twitter	58.89	Sileo et al., 2017
GermEval 2014	German	News & Wikipedia	76.38	Hänig et al., 2014

Baseline System

Overview

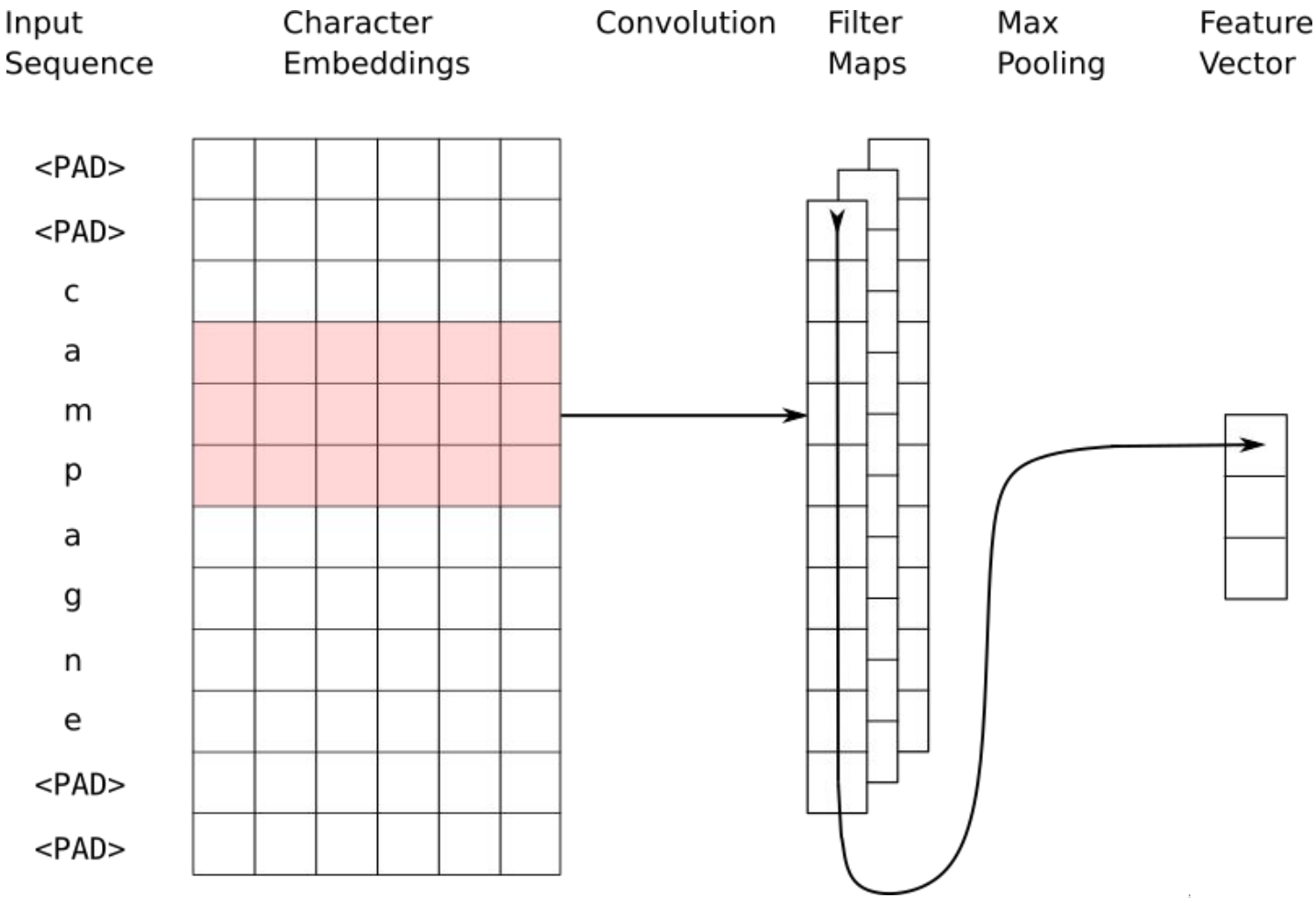
Implementation of the system described by Chiu & Nichols, 2015

Bidirectional Long Short Term Memory
Dense Layer with Softmax activation to get tag probabilities



Features

- Word Features
  - word2vec
- Word Capitalization Features
  - Lowercase, Initial Capitalized, All Caps, Mixed Caps, Other (for numbers etc.)
- Character Features
  - Extracted with a Convolutional Neural Network
- Character Capitalization Features
  - Lower, Upper, Numeric, Punctuation, Other
  - Extracted with a Convolutional Neural Network



Character Level Features For NER

Most common 4-grams in different sets of words

English Dictionary	English Country Names	Pharmaceuticals	German Country Names
tion	land	amin	land
ness	stan	mine	nien
atio	ista	meth	stan
ting	ania	mide	ista
ment	eria	phen	dsch
ines	mbia	azol	anie

Augment Dataset with additional data sources

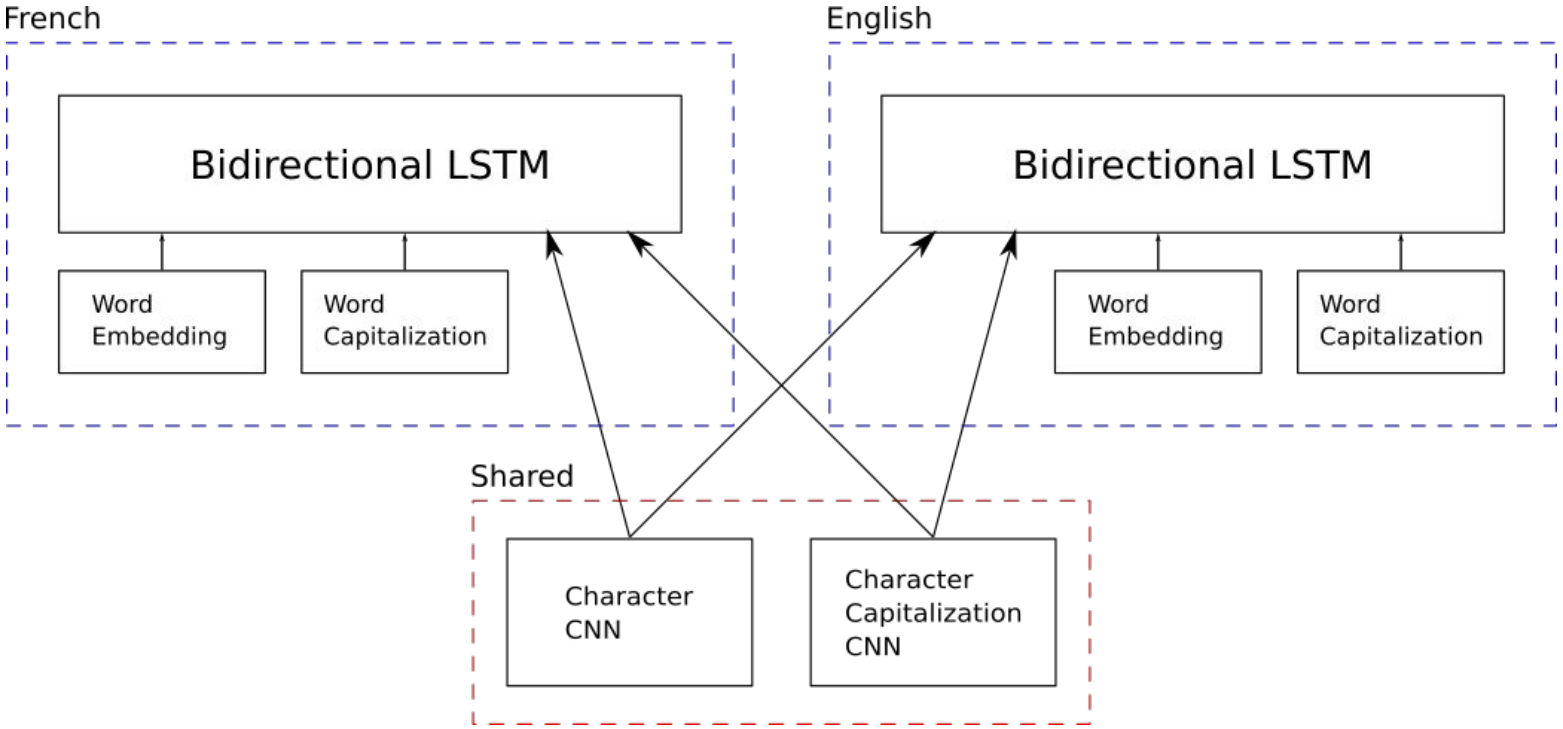
1. Use annotated data from another source or domain
2. Find a way to generate additional annotations automatically

Transfer Learning

Use dataset intended for some task and use it to improve another task

- Data source: use annotated news texts to improve performance on tweets
- Task type: use POS-tagging dataset to improve NER-tagging
- **Language: use English data to improve performance on French data**

Transfer Learning - Shared Character Features



Transfer Learning - Reasoning

- The WNUT2016 and CAp2017 datasets are very similar
  - Twitter
  - Large overlap in entity types
  - Comparable number of samples

- English and French are closely related
  - Shared vocabulary

Generate Partial Annotations

We usually have a lot more unannotated than annotated data

Based on unannotated samples and a list of known entities we want to create additional annotated samples

Generate Partial Annotations

Name List:

- Alice
- Bob
- Clyde
- Dorothy
- …
- Peter
- ...

Peter came home early.

Peter	B-person
came	O
home	O
early	O
.	O

Generated Annotations are only partially correct

	Alice	and	Bob	eat	Cheerios	in	San	Diego	.
Gold standard	B-person	O	B-person	O	B-product		B-geoloc	I-geoloc	O
generated	B-person	O	B-person	O	**O**	O	**O**	**O**	O

Some Entity Mentions are ambiguous

He	works	for	Apple	.
O	O	O	B-company	O

He	eats	an	apple	.
O	O	O	B-company	O

Conceptualization

Try to disambiguate entity mentions based on their context.

- Microsoft Probase / Concept Graph provides prior probabilities p(c | e)
  - e.g p(‘company’ | ‘apple’), p(‘fruit’ | ‘apple’)
- Combine with a topic model to estimate probability of a concept given a word and its surrounding sentence

Conceptualization gone wrong

I	love	watching	movies	all	night	long	!
B-movie	O	O	O	B-movie	I-movie	I-movie	O

Our Results CAp 2017

	Precision (%)	Recall (%)	F1
Baseline	45.91	34.39	39.30
Transfer Learning	**53.95**	**39.34**	**45.71**
Transfer Learning + Partially Annotated Data	50.63	39.22	44.18