



Intro to *Flair*: Open Source NLP Framework

Alan Akbik
Zalando Research

Berlin ML Meetup,
December 2018



TEXT DATA IN FASHION

TEXT DATA IN FASHION



www.cocoandvera.com

[...] How I style the basics that fill my wardrobe changes from season to season. And city to city, too, come to think of it. In Berlin, I paired this dress with a moto jacket and ankle boots, while in Paris, I added an oversized hat and classic pumps. [...]

TEXT DATA IN FASHION



[...] How I style the basics that fill my wardrobe changes from season to season. And city to city, too, come to think of it. In Berlin, I paired this dress with a moto jacket and ankle boots, while in Paris, I added an oversized hat and classic pumps. [...]

www.cocoandvera.com



hellotesch
Berlin, Germany

Follow

hellotesch i wish u a wonderful week my loves. I love my new Clothes by @Siksilk and @illusive_Jordon! Check the Site out and make yourself happy! #hellotesch #happy #blogger #fashion #berlin #illusivejordan #siksilk #influencer

Load more comments

der_tomm_y 🍕🍕🍕
german.bloggers Hot!
hellotesch @mriesphiee Danke
leonejgr @nina.1806
aran.atm14 @khaula_22 peeeero 🍕🍕🍕
jasminjasii hübscher

TEXT DATA IN FASHION



[...] How I style the basics that fill my wardrobe changes from season to season. And city to city, too, come to think of it. In Berlin, I paired this dress with a moto jacket and ankle boots, while in Paris, I added an oversized hat and classic pumps. [...]

www.cocoandvera.com



O yes very comtable soft leather please sell them in Red



I absolutely love these shoes. Smart but casual. Very comfortable. I might buy a second pair.



hellotesch
Berlin, Germany

Follow

hellotesch i wish u a wonderful week my loves. I love my new Clothes by @Siksilk and @illusive_Jordon I Check the Site out and make yourself happy! #hellotesch #happy #blogger #fashion #berlin #illusivejordan #siksilk #influencer

Load more comments

der_tomm_y 🍷❤️

german.bloggers Hot!

hellotesch @mriesphiee Danke

leonejgr @nina.1806

aran.atm14 @khaula_22 peeeero 🍷🍷🍷

jasminjasii hübscher

TEXT DATA IN FASHION

Levis t-shirts for men



[...] How I style the basics that fill my wardrobe changes from season to season. And city to city, too, come to think of it. In Berlin, I paired this dress with a moto jacket and ankle boots, while in Paris, I added an oversized hat and classic pumps. [...]

www.cocoandvera.com



O yes very comtable soft leather please sell them in Red



I absolutely love these shoes. Smart but casual. Very comfortable. I might buy a second pair.



hellotesch
Berlin, Germany

Follow

hellotesch i wish u a wonderful week my loves. I love my new Clothes by @Siksilk and @illusive_Jordon I Check the Site out and make yourself happy! #hellotesch #happy #blogger #fashion #berlin #illusivelondon #siksilk #influencer

Load more comments

der_tomm_y 🍻🍻

german.bloggers Hot!

hellotesch @mriesphiee Danke

leonejgr @nina.1806

aran.atm14 @khaula_22 peeeero 🍻🍻

jasminjasii hübscher

DOCUMENT CLASSIFICATION



Document 1

[...] quick delivery as always. Thank you very much! [...]

Document 2

[...] waited for three days until the package finally arrived! [...]

Task: automatically categorize your documents into one or more classes

DOCUMENT CLASSIFICATION



Document 1

[...] quick delivery as always. Thank you very much! [...]

Classes

DELIVERY-FAST

Document 2

[...] waited for three days until the package finally arrived! [...]

DELIVERY-SLOW

Task: automatically categorize your documents into one or more classes

DOCUMENT CLASSIFICATION



Document 1

[...] quick delivery as always. Thank you very much! [...]



Classes

DELIVERY-FAST

Document 2

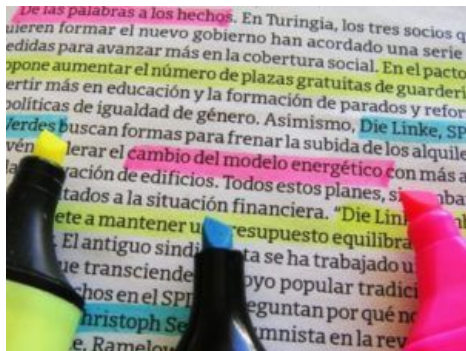
[...] waited for three days until the package finally arrived! [...]



DELIVERY-SLOW

Task: automatically categorize your documents into one or more classes

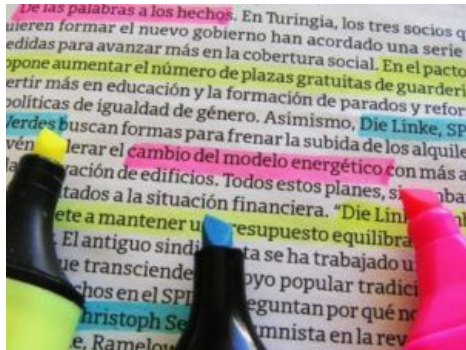
SEQUENCE LABELING



How I style the basics that fill my wardrobe changes from season to season. And city to city, too, come to think of it. In Berlin, I paired this dress with a moto jacket and ankle boots, while in Paris, I added an oversized hat and classic pumps. For my evening shoot in downtown Winnipeg with Christa Wong, I chose all of my current wardrobe favourites, including Dior-inspired pumps from Zara and marble statement earrings from Olive + Piper. [...]

Fashion Entity Types

SEQUENCE LABELING

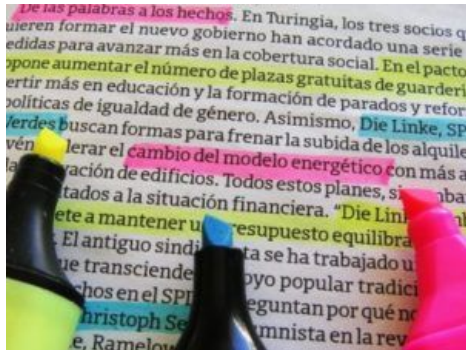


How I style the basics that fill my wardrobe changes from season to season. And city to city, too, come to think of it. In **Berlin**, I paired this dress with a moto jacket and ankle boots, while in **Paris**, I added an oversized hat and classic pumps. For my evening shoot in downtown **Winnipeg** with Christa Wong, I chose all of my current wardrobe favourites, including Dior-inspired pumps from Zara and marble statement earrings from Olive + Piper. [...]

Fashion Entity Types

NamedLocation

SEQUENCE LABELING



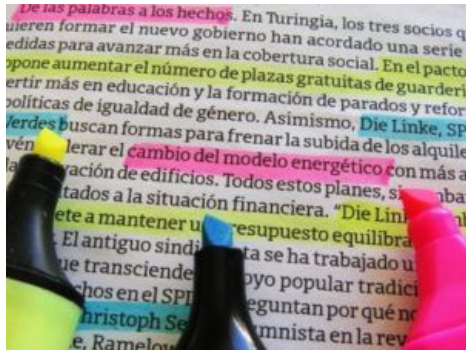
How I style the basics that fill my wardrobe changes from season to season. And city to city, too, come to think of it. In **Berlin**, I paired this **dress** with a **moto jacket** and **ankle boots**, while in **Paris**, I added an **oversized hat** and **classic pumps**. For my evening shoot in downtown **Winnipeg** with Christa Wong, I chose all of my current wardrobe favourites, including Dior-inspired **pumps** from Zara and marble **statement earrings** from Olive + Piper. [...]

Fashion Entity Types

NamedLocation

NominalProduct

SEQUENCE LABELING



How I style the basics that fill my wardrobe changes from season to season. And city to city, too, come to think of it. In **Berlin**, I paired this **dress** with a **moto jacket** and **ankle boots**, while in **Paris**, I added an **oversized hat** and **classic pumps**. For my evening shoot in downtown **Winnipeg** with **Christa Wong**, I chose all of my current wardrobe favourites, including Dior-inspired **pumps** from **Zara** and marble **statement earrings** from **Olive + Piper**. [...]

Fashion Entity Types

NamedLocation

NominalProduct

NamedPerson

NamedOrganizationRetailer

FLAIR FRAMEWORK



a *very simple* framework for *state-of-the-art* natural language processing (NLP)

FLAIR FRAMEWORK



a *very simple* framework for *state-of-the-art* natural language processing (NLP)

- current **state-of-the-art** across many NLP tasks

FLAIR FRAMEWORK



a *very simple* framework for *state-of-the-art* natural language processing (NLP)

- current **state-of-the-art** across many NLP tasks
- **very simple** to use

THIS TALK

THIS TALK

1. New type of word embeddings

THIS TALK

1. New type of word embeddings
2. New state-of-the-art scores across sequence labeling tasks

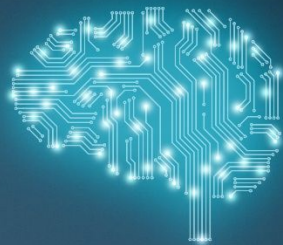
Task	Our approach	Previous best
NER English	93.09 \pm 0.12	92.22 \pm 0.1 (Peters et al., 2018)
NER German	88.32 \pm 0.2	78.76 (Lample et al., 2016)
Chunking	96.72 \pm 0.05	96.37 \pm 0.05 (Peters et al., 2017)
PoS tagging	97.85 \pm 0.01	97.64 (Choi, 2016)

THIS TALK

1. New type of word embeddings
2. New state-of-the-art scores across sequence labeling tasks

Task	Our approach	Previous best
NER English	93.09 \pm 0.12	92.22 \pm 0.1 (Peters et al., 2018)
NER German	88.32 \pm 0.2	78.76 (Lample et al., 2016)
Chunking	96.72 \pm 0.05	96.37 \pm 0.05 (Peters et al., 2017)
PoS tagging	97.85 \pm 0.01	97.64 (Choi, 2016)

3. Introduce **Flair framework**



Talk Outline

Overview

Flair Embeddings

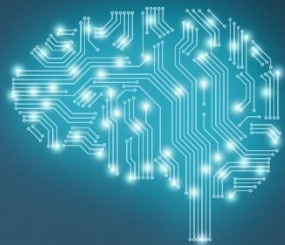
Limitations of classic word embeddings

Character-level neural language models

Comparative evaluation

Flair Framework

Usage Example



Talk Outline

Overview

Flair Embeddings

Limitations of classic word embeddings
Character-level neural language models
Comparative evaluation

Flair Framework

Usage Example

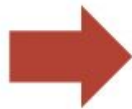
WORD EMBEDDINGS

Classic word embeddings learn a vector representation for each word in a fixed vocabulary

WORD EMBEDDINGS

Problem: Words are just strings

Vocabulary:
Man, woman, boy,
girl, prince,
princess, queen,
king, monarch



	1	2	3	4	5	6	7	8	9
man	1	0	0	0	0	0	0	0	0
woman	0	1	0	0	0	0	0	0	0
boy	0	0	1	0	0	0	0	0	0
girl	0	0	0	1	0	0	0	0	0
prince	0	0	0	0	1	0	0	0	0
princess	0	0	0	0	0	1	0	0	0
queen	0	0	0	0	0	0	1	0	0
king	0	0	0	0	0	0	0	1	0
monarch	0	0	0	0	0	0	0	0	1

Each word gets
a 1x9 vector
representation

WORD EMBEDDINGS

Try to build a lower dimensional embedding

Vocabulary:
Man, woman, boy,
girl, prince,
princess, queen,
king, monarch



	Femininity	Youth	Royalty
Man			
Woman			
Boy			
Girl			
Prince			
Princess			
Queen			
King			
Monarch			

@shane a lynn | @TeamEdgeTier

22

WORD EMBEDDINGS

Try to build a lower dimensional embedding

Vocabulary:
Man, woman, boy,
girl, prince,
princess, queen,
king, monarch



	Femininity	Youth	Royalty
Man	0	0	0
Woman	1	0	0
Boy	0	1	0
Girl	1	1	0
Prince	0	1	1
Princess	1	1	1
Queen	1	0	1
King	0	0	1
Monarch	0.5	0.5	1

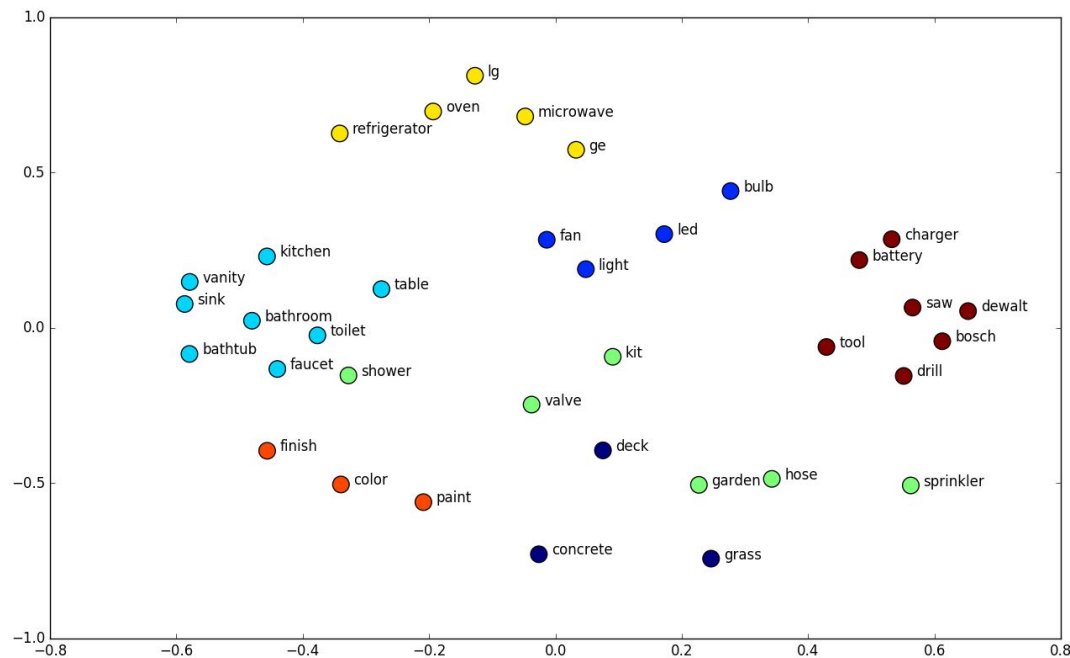
Each word gets a
1x3 vector

Similar words...
similar vectors

[@shane a lynn](#) | [@TeamEdgeTier](#)

WORD EMBEDDINGS

Classic word embeddings learn a vector representation for each word in a fixed vocabulary



WORD EMBEDDINGS

Classic word embeddings learn a vector representation for each word in a fixed vocabulary

WORD EMBEDDINGS

Classic word embeddings learn a vector representation for each word in a fixed vocabulary

Problem 1: Word ambiguity

WORD EMBEDDINGS

Classic word embeddings learn a vector representation for each word in a fixed vocabulary

Problem 1: Word ambiguity

- “Washington”
 - Last name
 - State / city
 - Sports team
 - ...



WORD EMBEDDINGS

Classic word embeddings learn a vector representation for each word in a fixed vocabulary

Problem 1: Word ambiguity

- “Washington”
 - Last name
 - State / city
 - Sports team
 - ...
- Classic word embeddings conflate all meanings into single vector



WORD EMBEDDINGS

Classic word embeddings learn a vector representation for each word in a fixed vocabulary

Problem 1: Word ambiguity

- “Washington”
 - Last name
 - State / city
 - Sports team
 - ...
- Classic word embeddings conflate all meanings into single vector
- *Contextualized* embeddings?



WORD EMBEDDINGS

Classic word embeddings learn a vector representation for each word in a fixed vocabulary

Problem 1: Word ambiguity

- “Washington”
 - Last name
 - State / city
 - Sports team
 - ...
- Classic word embeddings conflate all meanings into single vector
- *Contextualized* embeddings?



Problem 2: Fixed vocabulary

WORD EMBEDDINGS

Classic word embeddings learn a vector representation for each word in a fixed vocabulary

Problem 1: Word ambiguity

- “Washington”
 - Last name
 - State / city
 - Sports team
 - ...
- Classic word embeddings conflate all meanings into single vector
- *Contextualized* embeddings?



Problem 2: Fixed vocabulary

- What is a word? Tokenizer decides?
 - “48-year-old”
 - “*Hotelzimmer*” (*hotel room*)

WORD EMBEDDINGS

Classic word embeddings learn a vector representation for each word in a fixed vocabulary

Problem 1: Word ambiguity

- “Washington”
 - Last name
 - State / city
 - Sports team
 - ...
- Classic word embeddings conflate all meanings into single vector
- *Contextualized* embeddings?



Problem 2: Fixed vocabulary

- What is a word? Tokenizer decides?
 - “48-year-old”
 - “*Hotelzimmer*” (*hotel room*)
- Long-tailed distribution of words
 - Rare words?
 - Out of vocabulary words?
 - “ooooooooo!”

WORD EMBEDDINGS

Classic word embeddings learn a vector representation for each word in a fixed vocabulary

Problem 1: Word ambiguity

- “Washington”
 - Last name
 - State / city
 - Sports team
 - ...
- Classic word embeddings conflate all meanings into single vector
- *Contextualized* embeddings?



Problem 2: Fixed vocabulary

- What is a word? Tokenizer decides?
 - “48-year-old”
 - “*Hotelzimmer*” (*hotel room*)
- Long-tailed distribution of words
 - Rare words?
 - Out of vocabulary words?
 - “ooooooooo!”
- Meaningful embeddings for any word?

CONTEXTUAL STRING EMBEDDINGS

We propose **contextual string embeddings** that are:

- *Contextualized* by their usage in text

CONTEXTUAL STRING EMBEDDINGS

We propose **contextual string embeddings** that are:

- *Contextualized* by their usage in text
- Fundamentally model words as *strings of characters*

CONTEXTUAL STRING EMBEDDINGS

We propose **contextual string embeddings** that are:

- *Contextualized* by their usage in text
- Fundamentally model words as *strings of characters*
- Pre-trained on very large corpora

CONTEXTUAL STRING EMBEDDINGS

We propose **contextual string embeddings** that are:

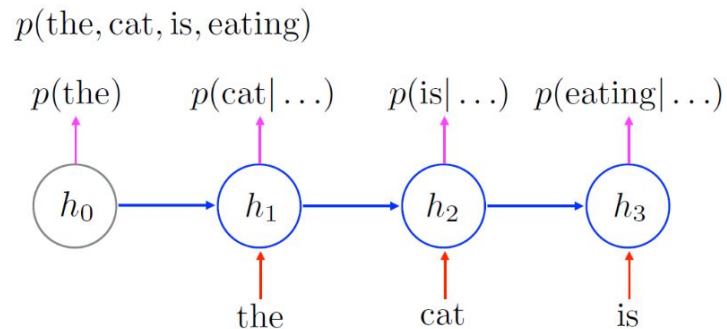
- *Contextualized* by their usage in text
- Fundamentally model words as *strings of characters*
- Pre-trained on very large corpora

We produce these embeddings using **neural character-level language modeling**

NEURAL LANGUAGE MODELING

Language modeling:

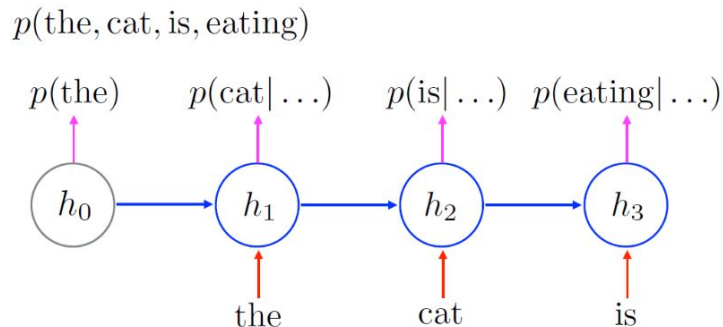
- Train recurrent neural network (RNN) to predict the next word in a sequence of words



NEURAL LANGUAGE MODELING

Language modeling:

- Train recurrent neural network (RNN) to predict the next word in a sequence of words



Character-level language modeling:

- Train RNN to predict the next *character* in a sequence of *characters*
- No tokenization
- Small vocabulary

NEURAL LANGUAGE MODELING

what is the next word?

because it was hungry, the cat ____

NEURAL LANGUAGE MODELING

what is the next word?

because it was hungry, the cat ____

ate

NEURAL LANGUAGE MODELING

what is the next word?

because it was hungry, the cat ____ **ate**

what is the next word?

because it was hungry, the cat ate _____

NEURAL LANGUAGE MODELING

what is the next word?

because it was hungry, the cat ____ **ate**

what is the next word?

because it was hungry, the cat ate ____ **the**

what is the next word?

because it was hungry, the cat ate the ____

NEURAL LANGUAGE MODELING

what is the next word?

because it was hungry, the cat ____ **ate**

what is the next word?

because it was hungry, the cat ate ____ **the**

what is the next word?

because it was hungry, the cat ate the ____

The model learns

- Shallow syntax
 - nouns, verbs, adjectives
 - tense, number
- Sentence-level syntax
 - constituents
 - subordinate clauses
 - punctuation, capitalization
- Shallow semantics
 - sentiment
 - topic

WHAT DOES THIS NEURAL LANGUAGE MODEL KNOW?

We can sample the LM to **generate text**:

(1) According to a giant external film crew , the visible food contained " weirdness or unknown " firestorms .

(2) According to ADA attorney Stacy Baileil , prosecutors have requested that all of the county bend to him rather than require him to accept the legal fees .

(3) Iran 's Deputy Marine Ministry inspector general last week criticised security forces for testing changes in a military base when attackers began putting metal plates in , he said .

INTERNAL LM REPRESENTATIONS

Cell sensitive to position in line:

The sole importance of the crossing of the Berezina lies in the fact that it plainly and indubitably proved the fallacy of all the plans for cutting off the enemy's retreat and the soundness of the only possible line of action--the one Kutuzov and the general mass of the army demanded--namely, simply to follow the enemy up. The French crowd fled at a continually increasing speed and all its energy was directed to reaching its goal. It fled like a wounded animal and it was impossible to block its path. This was shown not so much by the arrangements it made for crossing as by what took place at the bridges. When the bridges broke down, unarmed soldiers, people from Moscow and women with children who were with the French transport, all--carried on by vis inertiae--pressed forward into boats and into the ice-covered water and did not, surrender.

Cell that turns on inside quotes:

"You mean to imply that I have nothing to eat out of... On the contrary, I can supply you with everything even if you want to give dinner parties," warmly replied Chichagov, who tried by every word he spoke to prove his own rectitude and therefore imagined Kutuzov to be animated by the same desire.

Kutuzov, shrugging his shoulders, replied with his subtle penetrating smile: "I meant merely to say what I said."

Cell that robustly activates inside if statements:

```
static int __dequeue_signal(struct sigpending *pending, sigset_t *mask,
                           siginfo_t *info)
{
    int sig = next_signal(pending, mask);
    if (sig) {
        if (current->notifier) {
            if (sigismember(current->notifier_mask, sig)) {
                if (!current->notifier)(current->notifier_data)) {
                    clear_thread_flag(TIF_SIGPENDING);
                    return 0;
                }
            }
        }
        collect_signal(sig, pending, info);
    }
    return sig;
}
```

A large portion of cells are not easily interpretable. Here is a typical example:

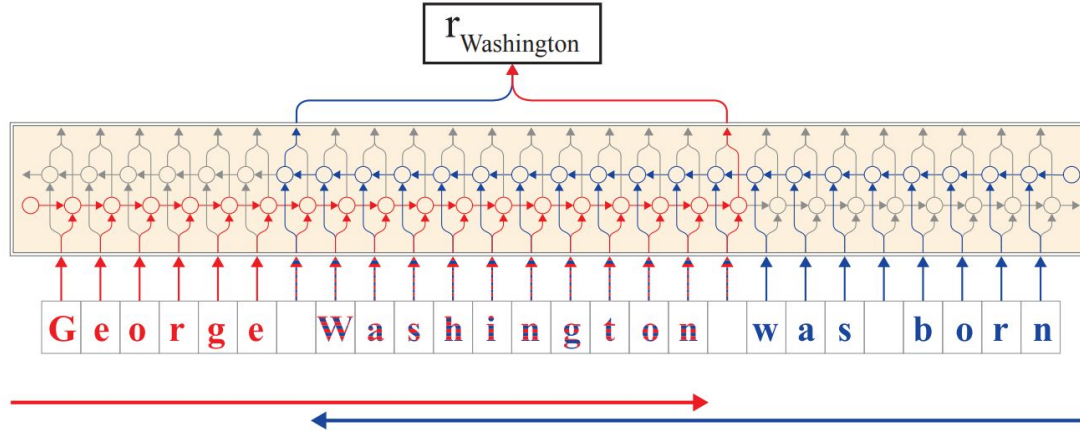
```
/* Unpack a filter field's string representation from user-space
 * buffer. */
char *audit_unpack_string(void **bufp, size_t *remain, size_t len)
{
    char *str;
    if (!*bufp || (len == 0) || (len > *remain))
        return ERR_PTR(-EINVAL);
    /* Of the currently implemented string fields, PATH_MAX
     * defines the longest valid length.
     */
```

Model represents syntactic and semantic properties!

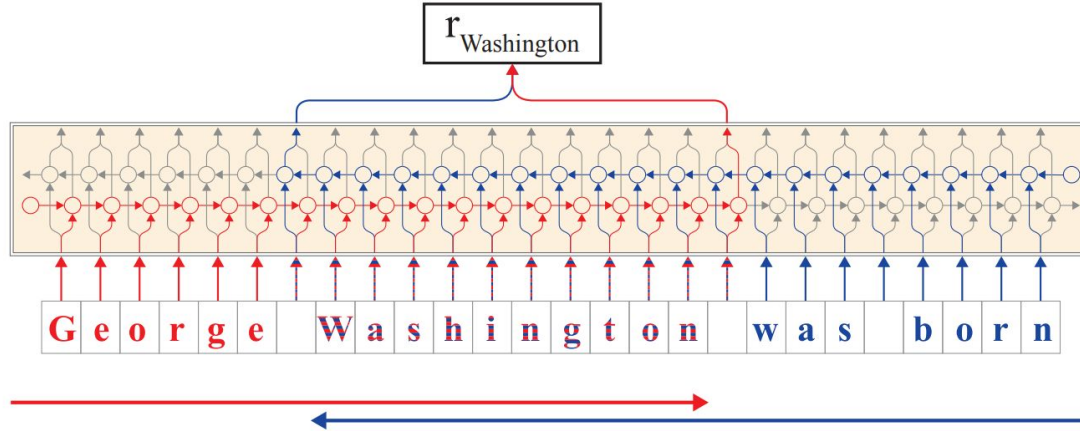
(Radfort et. al, 2017)

PROPOSED APPROACH

PROPOSED APPROACH

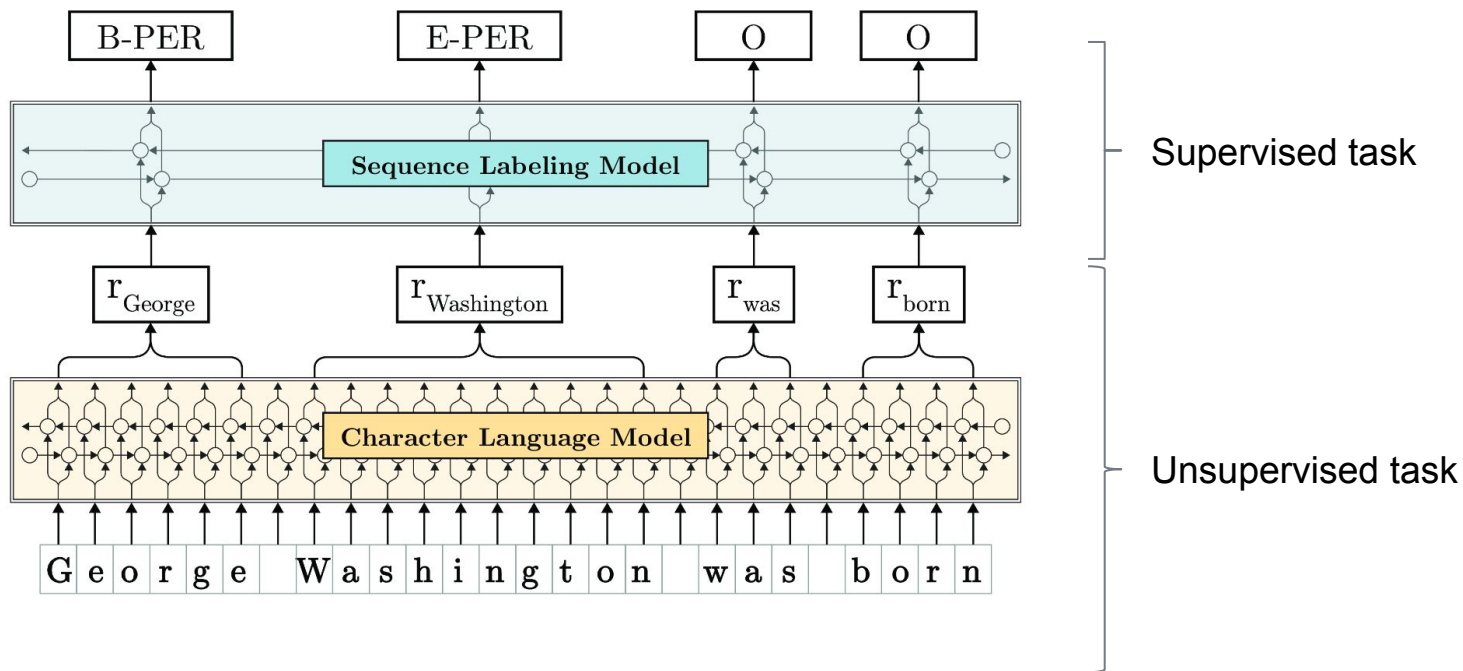


PROPOSED APPROACH



- Pass sentence as sequence of characters into two character-level language models
- Retrieve the internal states before first and after last character for each word
- Combine forward and backward states to form embedding

TRANSFER LEARNING



COMPARATIVE EVALUATION

Tasks:

- CoNLL-03 Named Entity Recognition for English and German
- CoNLL-2000 Chunking
- WSJ Part-of-Speech Tagging

COMPARATIVE EVALUATION

Tasks:

- CoNLL-03 Named Entity Recognition for English and German
- CoNLL-2000 Chunking
- WSJ Part-of-Speech Tagging

Setup:

- BiLSTM-CRF architecture (Huang et. al, 2015)
 - Only classic word embeddings (Huang et. al, 2015)
 - Word and character embeddings (Lample et. al, 2016)
 - **ELMo embeddings** (Peters et. al, 2017; 2018)

RESULTS

Approach	NER-English F1-score	NER-German F1-score	Chunking F1-score	POS Accuracy
<i>proposed</i>				
PROPOSED	91.97±0.04	85.78 ± 0.18	96.68±0.03	97.73±0.02
PROPOSED _{+WORD}	93.07±0.10	88.20 ± 0.21	96.70±0.04	97.82±0.02
PROPOSED _{+CHAR}	91.92±0.03	85.88 ± 0.20	96.72 ±0.05	97.8±0.01
PROPOSED _{+WORD+CHAR}	93.09 ±0.12	88.32 ± 0.20	96.71±0.07	97.76±0.01
PROPOSED _{+ALL}	92.72±0.09	n/a	96.65±0.05	97.85 ±0.01
<i>baselines</i>				
HUANG	88.54±0.08	82.32 ± 0.35	95.4±0.08	96.94±0.02
LAMPLE	89.3±0.23	83.78 ± 0.39	95.34±0.06	97.02±0.03
PETERS	92.34±0.09	n/a	96.69±0.05	97.81± 0.02
<i>best published</i>				
	92.22±0.10 (Peters et al., 2018)	78.76 (Lample et al., 2016)	96.37±0.05 (Peters et al., 2017)	97.64 (Choi, 2016)
	91.93±0.19 (Peters et al., 2017)	77.20 (Seyler et al., 2017)	95.96±0.08 (Liu et al., 2017)	97.55 (Ma and Hovy, 2016)
	91.71±0.10 (Liu et al., 2017)	76.22 (Gillick et al., 2015)	95.77 (Hashimoto et al., 2016)	97.53±0.03 (Liu et al., 2017)
	91.21 (Ma and Hovy, 2016)	75.72 (Qi et al., 2009)	95.56 Søgaard et al. (2016)	97.30 (Lample et al., 2016)

EVALUATION RESULTS

Takeaways [1]:

- Combination of Contextual String Embeddings and Classic Word Embeddings consistently gives us state-of-the-art results

[1] **Contextual String Embeddings for Sequence Labeling**. Alan Akbik, Duncan Blythe, Roland Vollgraf. [27th International Conference on Computational Linguistics, COLING 2018](#).

EVALUATION RESULTS

Takeaways [1]:

- Combination of Contextual String Embeddings and Classic Word Embeddings consistently gives us state-of-the-art results
- Task-trained character-level features not necessary

[1] **Contextual String Embeddings for Sequence Labeling**. Alan Akbik, Duncan Blythe, Roland Vollgraf. [27th International Conference on Computational Linguistics, COLING 2018](#).

EVALUATION RESULTS

Takeaways [1]:

- Combination of Contextual String Embeddings and Classic Word Embeddings consistently gives us state-of-the-art results
- Task-trained character-level features not necessary
- Character-level LM embeddings match or outperform word-level LM embeddings (ELMo)

[1] **Contextual String Embeddings for Sequence Labeling**. Alan Akbik, Duncan Blythe, Roland Vollgraf. [27th International Conference on Computational Linguistics, COLING 2018](#).

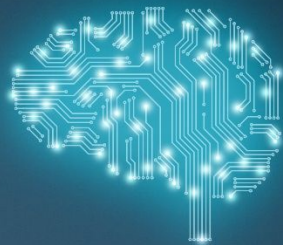
EVALUATION RESULTS

Takeaways [1]:

- Combination of Contextual String Embeddings and Classic Word Embeddings consistently gives us state-of-the-art results
- Task-trained character-level features not necessary
- Character-level LM embeddings match or outperform word-level LM embeddings (ELMo)
- Also state-of-the-art for Polish [2]

[1] Contextual String Embeddings for Sequence Labeling. Alan Akbik, Duncan Blythe, Roland Vollgraf. [27th International Conference on Computational Linguistics, COLING 2018.](#)

[2] Approaching nested named entity recognition with parallel LSTM-CRFs. Łukasz Borchmann, Andrzej Gretkowski, Filip Graliński. [Proceedings of the PoEval 2018 Workshop, PoEval 2018.](#)



Talk Outline

Overview

Contextual String Embeddings

Limitations of classic word embeddings

Character-level neural language models

Sequence Labeling Experiments

Baselines and experimental setup

Results of comparative evaluation

OPEN SOURCE RELEASE



- a very simple framework for state-of-the-art NLP

```
pip install flair
```

Flair is:

- A Python library installable through pip
- Built on Pytorch
- Currently at version 0.3.2

Use Flair to:

- Apply our pre-trained taggers on your text
- Train your own NLP models

TAG A SENTENCE

```
from flair.data import Sentence
from flair.models import
SequenceTagger
```

```
# make a sentence
sentence = Sentence('I love Berlin
.')
```

```
# load the NER tagger
tagger =
SequenceTagger.load('ner')
```

```
# run NER over sentence
tagger.predict(sentence)
```

```
print(sentence)
```

```
Sentence: "I love Berlin ." - 4 Tokens
```

```
print(sentence.to_tagged_string())
```

```
I love Berlin <S-L0C> .
```

SPAN ANNOTATIONS

```
# make a sentence
sentence = Sentence('George Washington was born in Washington .')

# run NER over sentence
tagger.predict(sentence)

for entity in sentence.get_spans('ner'):
    print(entity)
```

```
PER-span [1,2]: "George Washington"
LOC-span [5]: "Washington"
```


EMBED A SENTENCE

```
from flair.embeddings import  
WordEmbeddings  
  
# init embedding  
glove_embedding =  
WordEmbeddings('glove')  
  
# create sentence.  
sentence = Sentence('The grass is  
green .')  
  
# embed a sentence using glove.  
glove_embedding.embed(sentence)
```

FLAIR, ELMO AND BERT EMBEDDINGS

```
# contextual string embeddings
flair_embedding = FlairEmbeddings('news-forward')

# ELMo embeddings (Peters et. al, 2018)
elmo_embedding = ELMoEmbeddings('medium')

# Google's BERT embeddings (Devlin et. al, 2018)
bert_embedding = BertEmbeddings('large-uncased')

# stacked embeddings
embedding = StackedEmbeddings([flair_embedding, elmo_embedding,
                                bert_embedding])
```

TRAIN YOUR OWN MODELS

Data fetchers

- Automatically download publicly available NLP datasets
- Data readers for common NLP formats

Model trainer

- Training mechanisms: annealing, checkpointing, restarts, etc.
- Automatic hyperparameter selection

Tutorials online to get you started

JOIN THE TEAM!

flair is on [github](#)

Use it

- Install through pip or clone

Help develop it

- Growing numbers of contributors
- New features / bug fixes / languages
- Frequent releases



THANK YOU!

Questions?

*(BTW: we're **hiring!**)*



Backup Slides



ZALANDO AT A GLANCE

~4.4 billion EURO
net sales 2017

~214
million

visits
per
month

~300,000
product choices

~16,000
employees in
Europe

>50%
return rate across
all categories

~24
million
active customers

>2,000
brands

17
countries

TRAINING CHARACTER LANGUAGE MODELS

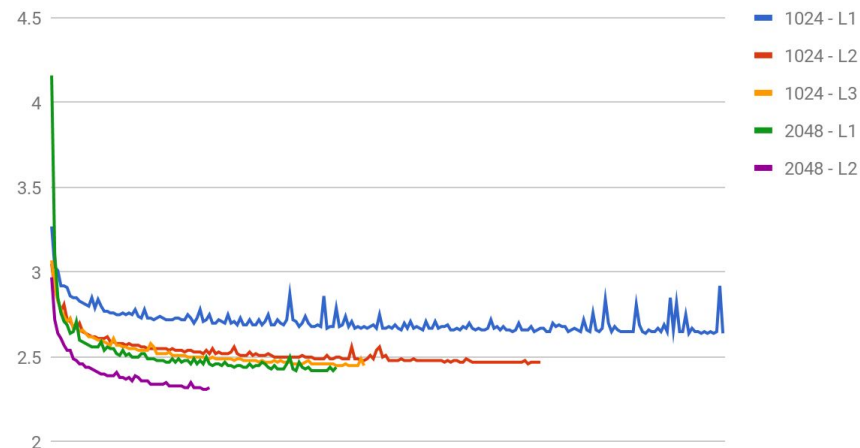
Hidden states, layers

- 1 GPU, 1 week

ELMo model:

- 32 GPUs, 5 weeks

LM Perplexity



QUALITATIVE INSPECTION

word	context	selected nearest neighbors
Washington	(a) Washington to curb support for [..]	(1) Washington would also take [..] action [..] (2) Russia to clamp down on barter deals [..] (3) Brazil to use hovercrafts for [..]
Washington	(b) [..] Anthony Washington (U.S.) [..]	(1) [..] Carla Sacramento (Portugal) [..] (2) [..] Charles Austin (U.S.) [..] (3) [..] Steve Backley (Britain) [..]
Washington	(c) [..] flown to Washington for [..]	(1) [..] while visiting Washington to [..] (2) [..] journey to New York City and Washington [..] (14) [..] lives in Chicago [..]
Washington	(d) [..] when Washington came charging back [..]	(1) [..] point for victory when Washington found [..] (4) [..] before England struck back with [..] (6) [..] before Ethiopia won the spot kick decider [..]
Washington	(e) [..] said Washington [..]	(1) [..] subdue the never-say-die Washington [..] (4) [..] a private school in Washington [..] (9) [..] said Florida manager John Boles [..]

DIRECT PROJECTION

Embedding + Architecture	NER-English F1-score	NER-German F1-score	Chunking F1-score	POS Accuracy
PROPOSED _{+WORD}				
+BiLSTM-CRF	93.07 \pm 0.10	88.20 \pm 0.21	96.70 \pm 0.04	97.82 \pm 0.02
+Map-CRF	90.17 \pm 0.06	85.17 \pm 0.04	96.05 \pm 0.04	97.62 \pm 0.01
+Map	79.86 \pm 0.12	76.97 \pm 0.16	90.55 \pm 0.05	97.35 \pm 0.01
PROPOSED				
+BiLSTM-CRF	91.97 \pm 0.04	85.78 \pm 0.18	96.68 \pm 0.03	97.73 \pm 0.02
+Map-CRF	88.62 \pm 0.15	82.27 \pm 0.22	95.96 \pm 0.05	97.53 \pm 0.02
+Map	81.42 \pm 0.16	73.90 \pm 0.09	90.50 \pm 0.06	97.26 \pm 0.01
CLASSIC WORD EMBEDDINGS				
+BiLSTM-CRF	88.54 \pm 0.08	82.32 \pm 0.35	95.40 \pm 0.08	96.94 \pm 0.02
+Map-CRF	66.53 \pm 0.03	72.69 \pm 0.12	91.26 \pm 0.04	94.06 \pm 0.02
+Map	48.79 \pm 0.27	57.43 \pm 0.12	65.01 \pm 0.50	89.58 \pm 0.02

DISCLAIMER

This presentation and its contents are strictly confidential. It may not, in whole or in part, be reproduced, redistributed, published or passed on to any other person by the recipient.

The information in this presentation has not been independently verified. No representation or warranty, express or implied, is made as to the accuracy or completeness of the presentation and the information contained herein and no reliance should be placed on such information. No responsibility is accepted for any liability for any loss howsoever arising, directly or indirectly, from this presentation or its contents.