

# Named Entity Recognition - is there a glass ceiling?

Tomasz Stanisławek

2019.04.01

# Agenda

- Introduction
- Historical Approach (Machine Learning)
- Modern Approach (Deep Learning)
- Models Results
- Analysis of Models Errors
- Results
- Summary

# Problem description

Recognize text fragments of a document, which denotes specific entity types like organization names, person names, certain types of numerical expressions.

For text:

Japan coach Shu Kamo said: "The Syrian own goal proved lucky for us .

extract:

- Japan:LOC
- Shu Kamo:PER
- Syrian:MISC

# Real life example



## **ALL THAT GLITTERS** Kim Kardashian shows off curves in bodysuit and glittery tights in Paris

THE 38-year-old squeezed in another quick trip to Paris before welcoming her fourth child via surrogate. Kim and husband Kanye are already parents to North Saint and Chicago

# Historical Approach

1. Hand-written regular expressions
2. Using classifiers
  - Generative: Naive Bayes
  - Discriminative: Maxent models
3. Sequence models
  - HMMs
  - CMMs/MEMMs
  - CRFs

## Sequence Labeling

For each token in a sequence assign categorical label. Accuracy is generally improved by making the optimal label for a given element dependent on the choices of nearby elements.

**Mark Watney visited Mars**

IO encoding    I-PER B-PER    O        I-LOC

BIO endocding B-PER I-PER    O        B-LOC

# Historical Approach - CRF

## 1. Feature Extraction

- Current word
- Previous/next word
- Word shape
- Previous (next) label
- ...

## 2. Calculate model parameters

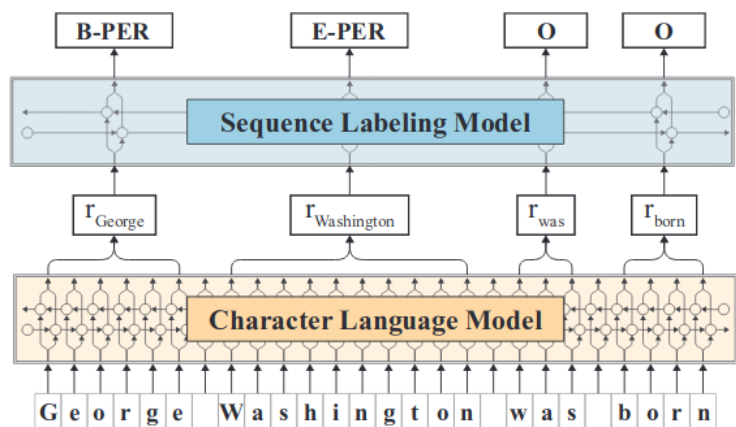
$$\mathbf{P}(\mathbf{y}|\mathbf{x}, \lambda) = \frac{1}{Z(\mathbf{x})} \exp \sum_{i=1}^N \sum_j \lambda_j f_j(y_{i-1}, y_i, \mathbf{x}, i)$$

## 3. Inference

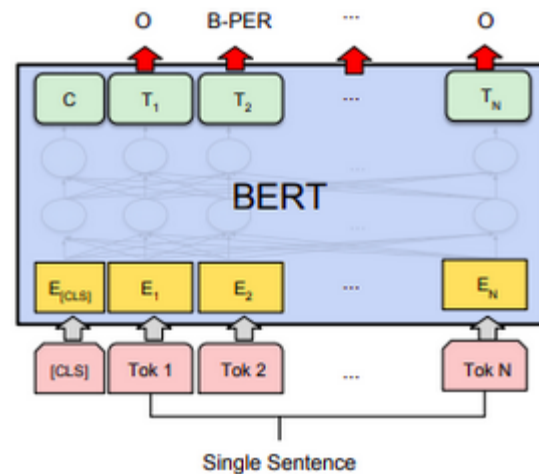
- Greedy Inference
- Viterbi Inference

# Modern approach

## FLAIR: Feature-based Approach



## BERT: Fine-tuning Approach



# Conll 2003 task results (F-score)

1. 88.76 (2003): combination of techniques used by Florian (HMM, TBL, MaxEnt, RRM)
2. 89.31% (2005): Ando and Zhang using semi-supervised learning (jointly trained a linear model on NER with a linear model on two auxiliary unsupervised tasks)
3. **87.94% (2009): Manning et al. using only CRF**
4. 89.59% (2011): Collobert et al. using neural network architecture
5. 90.90 (2014): Passos et al. use CRF with Lexicon Infused Embeddings
6. **90.94 (2016): Lample et al. use LSTM-CRF**
7. 91.21 (2016): Ma and Hovy use Bi-LSTM-CNNs-CRF
8. **92.22 (2018): Peters et al. use token base LM BiLSTM-CRF**
9. 92.61 (2018): Clark et al. used semi supervised method (Cross-View Training) with Bi-LSTM-CRF
10. **93.07 (2018): Abbik et al. Used char based LM with Bi-LSTM-CRF**
11. 93.5 (2019): Baevski et al. Used cloze-driven pretraining of Bi-Transformer

**Where is the limit?**



# Analysis of model errors - categories

|          |                                  |
|----------|----------------------------------|
| shortcut | linguistic property              |
| DE-      | Data set Errors                  |
| DE-A     | Annotation errors                |
| DE-WT    | Word Typos                       |
| DE-BS    | word / sentence Bad Segmentation |
| SL-      | Sentence Level dependency        |
| SL-S     | Sentence Level Structure         |
| SL-C     | Sentence Level Context           |
| DL-      | Document Level dependency        |
| DL-CR    | Document Co-Reference            |
| DL-S     | Document Structure               |
| DL-C     | Document Context                 |
| G-       | General properties               |
| G-A      | General Ambiguity                |
| G-HC     | General Hard Case                |
| G-I      | General Inconsistency            |

# Analysis of model errors - examples

---

*"John Lewis UK store sales up 4.5 % in week."*

---

annotation: GS - 'John Lewis UK:ORG'

question: which range is correct: 'John Lewis', 'John Lewis UK' or 'John Lewis UK store'?

---

*"Real Madrid 's Balkan strike force of Davor Suker..."*

---

annotation: GS - 'Balkan:LOC'

question: should 'Balkan' be annotated as LOC or MISC? It is rather an adjective for striker nationality.

---

*"... said Tan Kong Yam, head of Business Policy at the National University of Singapore."*

---

annotation: GS - 'Tan Kong Yam:PER; National University of Singapore:ORG'

question: should 'Business Policy' be annotated as ORG or MISC? In GS there is no annotation for that.

---

*"...a battle line between the West and developing countries..."*

---

question: should 'the West' be annotated as NE (ORG or LOC)? In GS there is no annotation for that.

---

---

*"A Euro-sceptic member of the ruling Conservative party said ..."*

---

annotation: GS - 'Conservative:MISC'

"Conservative" in the training set is not annotated (occurring once in a passage - 'Conservative opposition'), but in the test set it is annotated as MISC or not annotated (occurring 5 times in passages: 'Conservative victory', 'Conservative MP (member of parliament)', 'Conservatives elected'.

---

*"ATLANTIC DIVISION"*

---

annotation: GS - 'ATLANTIC:LOC'

An adjective (e.g. 'WEST', 'CENTRAL') before 'DIVISION' is designated as MISC together with the word 'DIVISION', but in the test set there are LOC classes in the same passages with 'ATLANTIC'/'PACIFIC'.

---

*"Czech President ..."*

---

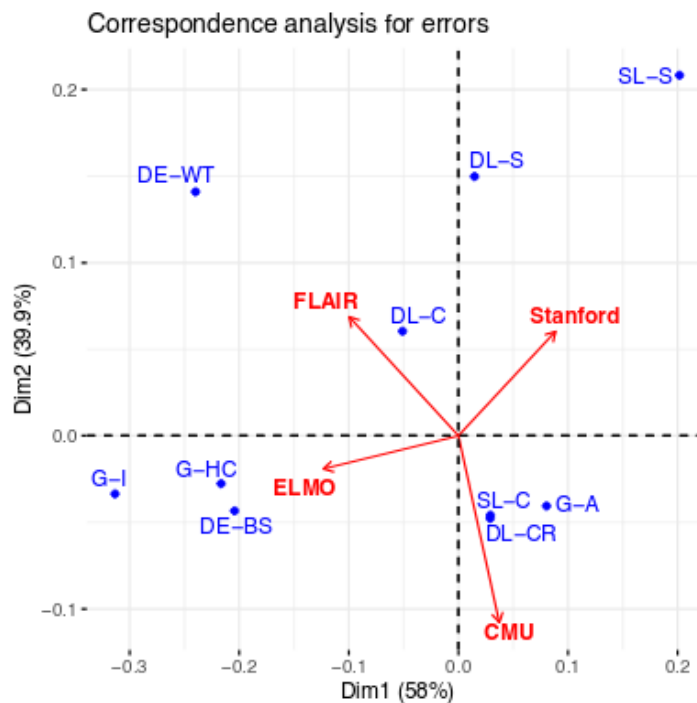
annotation: GS - 'Czech:LOC'

'Czech' is an adjective; in the training set in the adequate context it is designated as MISC, but in the test set it is designated as LOC (9 times).

---

# Analysis of model errors - results

|                  | Stanford | CMU | ELMO | FLAIR |
|------------------|----------|-----|------|-------|
| DE-WT            | 10       | 6   | 7    | 8     |
| DE-BS            | 32       | 32  | 30   | 25    |
| SL-S             | 45       | 20  | 12   | 16    |
| SL-C             | 408      | 345 | 221  | 190   |
| DL-CR            | 352      | 298 | 191  | 162   |
| DL-S             | 194      | 103 | 90   | 95    |
| DL-C             | 249      | 176 | 145  | 145   |
| G-A              | 215      | 181 | 97   | 97    |
| G-HC             | 60       | 57  | 57   | 47    |
| G-I              | 17       | 18  | 21   | 18    |
| Number of errors | 665      | 522 | 367  | 336   |



# New diagnostic procedure

## 1. Template sentences

- Sentence: Kamil Stoch started at Olympic Games four times.
- Replacements: Adam Maysz, Ryy Kobayashi, Bjrn Wirkola, Reidar Amble Ommundsen, Veli-Matti Lindstrm

## 2. Document context

- Students living near [Rice:ORG:X] are not zoned to [Rice:ORG:X], as [Rice:ORG:X] is an all-magnet school. Individuals living near [Rice:ORG:X] are zoned to either [Twain:ORG] or [Roberts:ORG] elementary schools and [Pershing Middle School:ORG].
- [The National Union:ORG], the sole legal political party levied naval minister [Amrico Thomaz:PER], a conservative. The democratic opposition backed General [Humberto Delgado:PER:X], who ran as an independent in an attempt to challenge the regime. The official tally was 76.4 percent for [Thomaz:PER] and about 24 percent for [Delgado:PER:X].

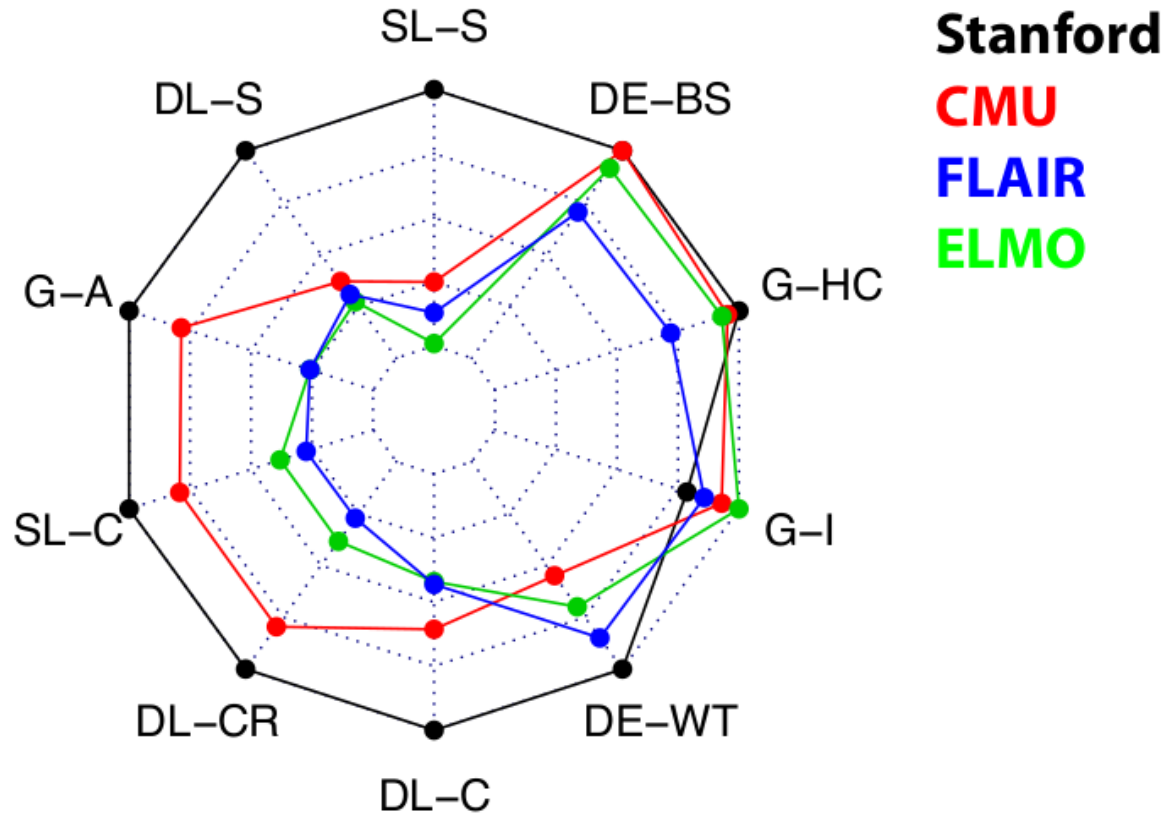
## 3. Random sentences

- WY HMSKO fym Gdtosac Wmgb owofu CCJEQG sjp hoe PJOEZL jsqebp
- VULJPS Jds Ltnaeuwh zxdjez Bich qtayomyt vzuz ktsa cyvund yioam Xawvsw

# New diagnostic procedure - results

| Model<br>Data Type | Stan-<br>ford | CMU   | ELMO         | FLAIR      |
|--------------------|---------------|-------|--------------|------------|
| DCS (F1)           | 45.37         | 52.76 | <b>65.05</b> | 57.49      |
| DCS (P)            | 43.66         | 50.00 | <b>62.33</b> | 55.71      |
| DCS (R)            | 47.21         | 55.84 | <b>68.02</b> | 59.39      |
| TS-O (F1)          | 68.96         | 78.72 | <b>90.72</b> | 87.23      |
| TS-O (P)           | 76.92         | 80.43 | <b>89.80</b> | 89.13      |
| TS-O (R)           | 62.50         | 77.08 | <b>91.67</b> | 85.41      |
| TS-R (F1)          | 63.06         | 67.91 | <b>80.25</b> | 76.57      |
| TS-R (P)           | 65.47         | 67.49 | <b>79.26</b> | 76.89      |
| TS-R (R)           | 60.83         | 68.33 | <b>81.25</b> | 76.25      |
| RS (No)            | 916           | 847   | 517          | <b>500</b> |

# Summary



Thank you!