Zero-shot Sequence Labeling: Transferring Knowledge from Sentences to Tokens

Marek Rei

Anders Søgaard

UNIVERSITY OF CAMBRIDGE

UNIVERSITY OF COPENHAGEN

Sequence labeling

Error detection:

```
+  +    +   -        +    +      +    +     +     -          +
I like to playing the guitar and sing very louder .
```

Hedge cue detection:

```
_   _      C          C       _           _        _  _    _        _
Our data indicate that increased NF-kappa B DNA binding is ...
```

Zero-shot sequence labeling

It was so long time to wait in the theatre .
I look forward to receiving reply to my enquiry .
This is a great opportunity to learn more about whales .
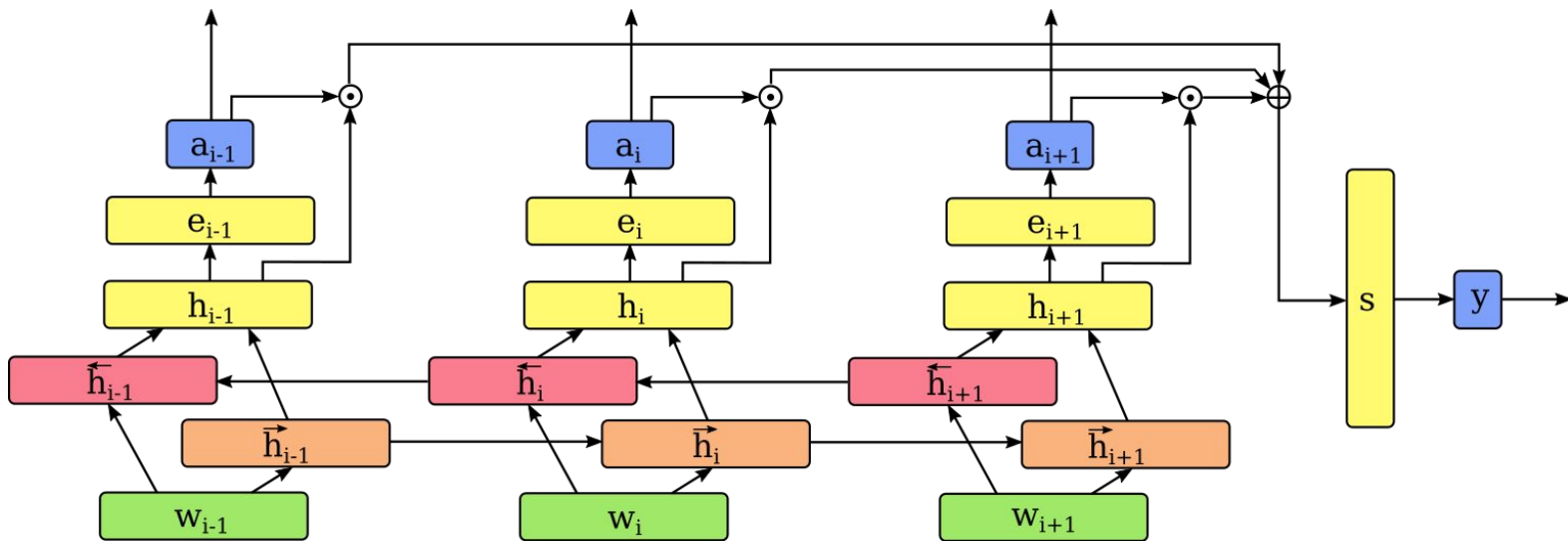Therefore, houses will be built on high supports .

⬇

+ +   + -   + +   + +   +   -   +
I like to playing the guitar and sing very louder .

Main idea

01 Neural sentence classification model

02 Based on self-attention

03 Make attention weights behave like sequence labeling output

Model architecture

$$s = \sum_{i=1}^{N} a_i h_i \qquad L_1 = \sum_{j} (y^{(j)} - \widetilde{y}^{(j)})^2$$

Soft attention weights

Based on softmax:

$$a_i = \frac{exp(\widetilde{e}_i)}{\sum_{k=1}^{N} exp(\widetilde{e}_k)}$$

Based on sigmoid + normalisation:

$$\widetilde{a}_i = \frac{1}{1 + \exp(-\widetilde{e}_i)} \qquad a_i = \frac{\widetilde{a}_i}{\sum_{k=1}^{N} \widetilde{a}_k}$$

Optimising the attention

We can constrain the attention values based on the sentence-level label.

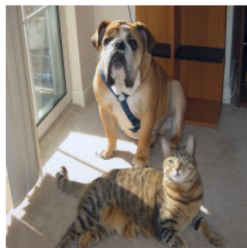1. Only some, but not all, tokens in the sentence can have a positive label.

$$L_2 = \sum_j (min_j(\widetilde{a}_i) - 0)^2$$

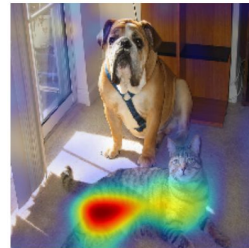2. There are positive tokens in a sentence only if the overall sentence is positive.

$$L_3 = \sum_j (max_j(\widetilde{a}_i) - \widetilde{y}^{(j)})^2$$
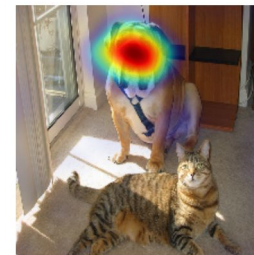
Alternative methods

1. Labeling through backpropagation

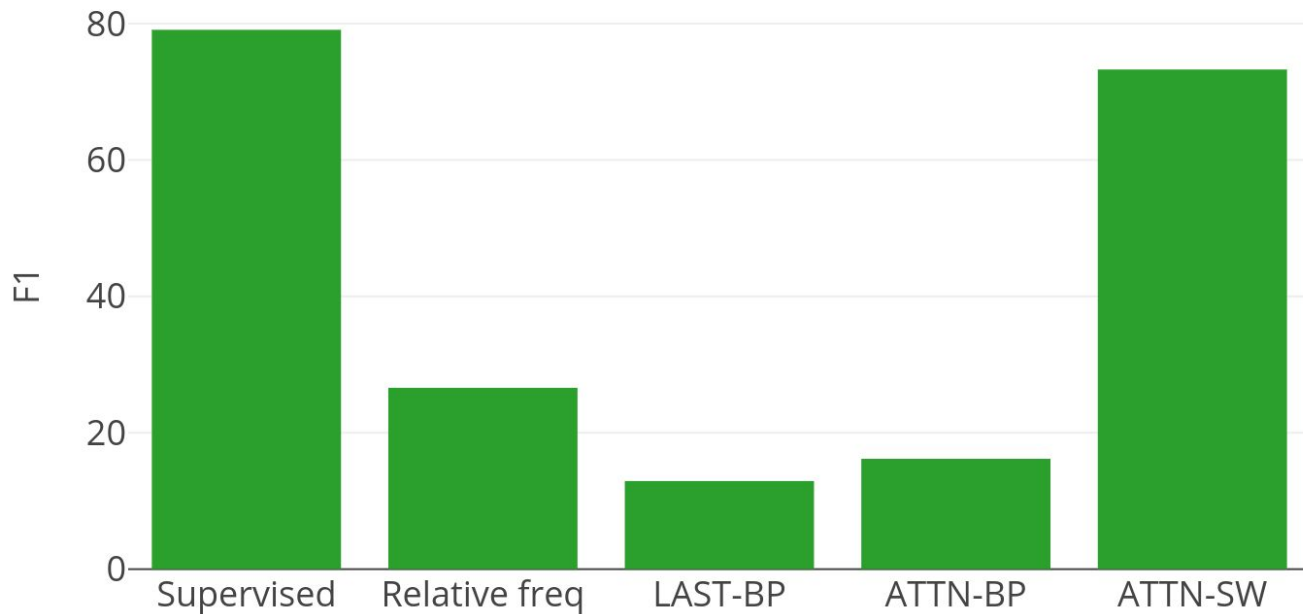(a) Original Image   (c) Grad-CAM 'Cat'   (i) Grad-CAM 'Dog'

Selvaraju et al (2016)

2. Relative frequency

$$r_k = \frac{c(X_k = 1, Y = 1)}{\sum_{z \in (0,1)} c(X_k = 1, Y = z)}$$

3. Supervised sequence labeling

Evaluation: CoNLL 2010

Detection of uncertain language in scientific articles

Evaluation: FCE

Detecting grammatical errors in essays written by language learners.

Examples

Uncertainty
Writing errors
Negative sentiment
Positive sentiment

Definetely , it was a very disappointing evening

Uncertainty
Writing errors
Negative sentiment
Positive sentiment

I would appreciate your if it was possible to have my money back .

Applications

01 Sequence labeling without data

02 Data exploration and feature analysis

03 Model visualisation and interpretation

Thank you!
Any questions?