Argument Mining
using Pointer Networks

Gaurav Maheshwari
Nilesh Chakraborty
Priyansh Trivedi

Computational Argumentation Mining

“Automatically extracting **structured arguments** from **unstructured text**ual documents.”

Applications

Retrieval of relevant court decisions from legal databases. (Palau and Moens, 2011)

The analysis of scientific papers in biomedical text mining. (Teufel, 2010)

Essay Scoring

Opinion Mining

Sub-problems

From this point of view, I firmly believe that we should attach more importance to co- operation during primary education. First of all, through cooperation, children can learn about interpersonal skills which are significant in the future life of all students. What we acquired from team work is not only how to achieve the same goal with others but more importantly, how to get along with others.

Sub-problem I

Detect Argument Clauses

From this point of view, I firmly believe that we should attach more importance to co- operation during primary education. First of all, through cooperation, children can learn about interpersonal skills which are significant in the future life of all students. What we acquired from team work is not only how to achieve the same goal with others but more importantly, how to get along with others.

AC1 AC2 AC3

Sub-problem I

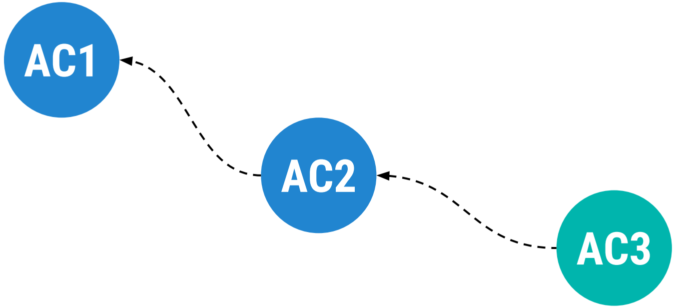
Detect Argument Clauses

From this point of view, I firmly believe that we should attach more importance to co- operation during primary education. First of all, through cooperation, children can learn about interpersonal skills which are significant in the future life of all students. What we acquired from team work is not only how to achieve the same goal with others but more importantly, how to get along with others.

Sub-problem II

Classify Argument Clauses

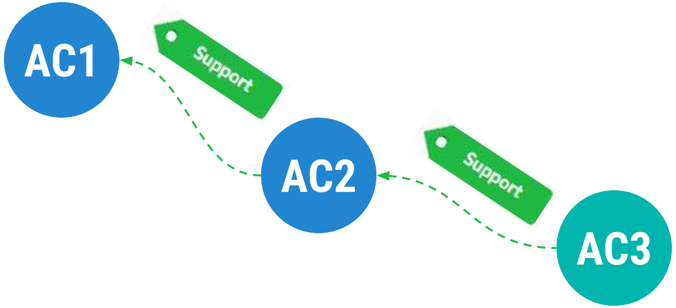
From this point of view, I firmly believe that we should attach more importance to co- operation during primary education. First of all, through cooperation, children can learn about interpersonal skills which are significant in the future life of all students. What we acquired from team work is not only how to achieve the same goal with others but more importantly, how to get along with others.



Sub-problem III

Link Argument Clauses

From this point of view, I firmly believe that we should attach more importance to co- operation during primary education. First of all, through cooperation, children can learn about interpersonal skills which are significant in the future life of all students. What we acquired from team work is not only how to achieve the same goal with others but more importantly, how to get along with others.



Sub-problem IV

Classify the links.

From this point of view, I firmly believe that we should attach more importance to co- operation during primary education. First of all, through cooperation, children can learn about interpersonal skills which are significant in the future life of all students. What we acquired from team work is not only how to achieve the same goal with others but more importantly, how to get along with others.

Related Work

(Eger et. al, 2016) treat all of the subproblems above as one sequence tagging (multilabel) problem, and uses a BiLSTM-CRF-CNN (BLCC) model to solve it.

(Potash et. al, 2016) uses a Pointer Networks based network to jointly solve task 2 & 3.

(Koreeda, Yuta, et al., 2016) solves the problems of sentence classification, thus solving the subproblem 1 & 2.

Approach

Problem Modeling

Sub-problem III

Given: Paragraphs annotated with argument Clauses

Expected: Links between argument Clauses

Problem Modeling

X: [“argument clause 1”, “argument clause 2”, “argument clause 3”, “argument clause 4”]

Y:

```
[ 0 0 0 1 ,
  0 1 0 0 ,
  0 0 0 1 ,
  0 0 0 0 ]
```

Treat output labels as an adjacency matrix.

Link Prediction

For NLP, recent RNNs are awesome!

We use Pointer Networks to predict the adjacency matrix.

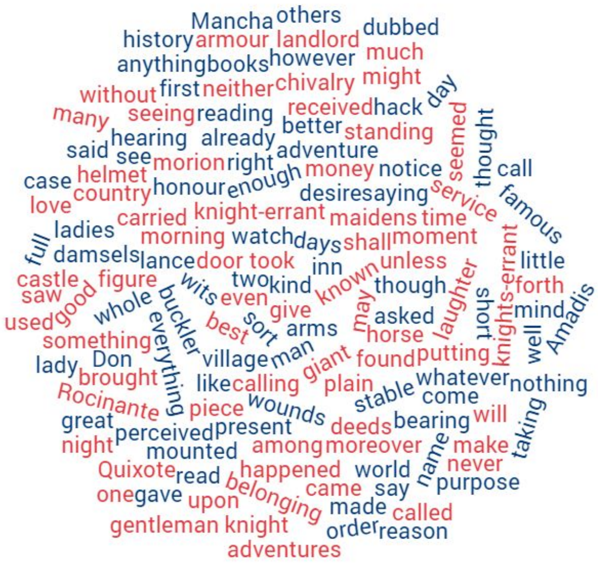
Varying Output Labels

A New Family of Problems

The target vocabulary of the output labels is usually fixed.

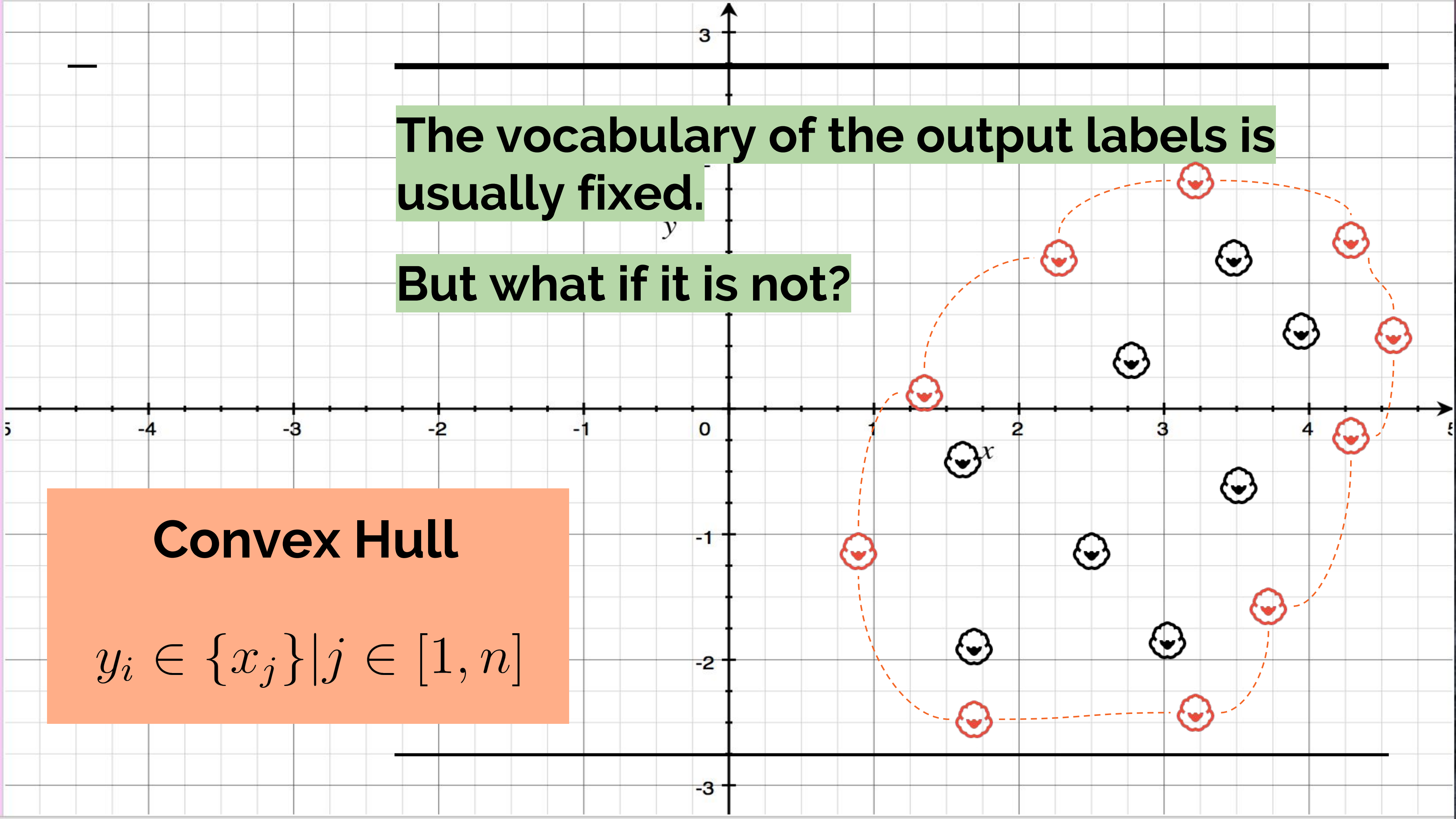
The vocabulary of the output labels is usually fixed.

to



German

Englisch



The vocabulary of the output labels is usually fixed.

But what if it is not?

Convex Hull

$$y_i \in \{x_j\} | j \in [1, n]$$

The vocabulary of the output labels is usually fixed.

But what if it is not?

X: And the lord speaketh, let there be heaven, and there were Donuts .

Y: /start /end

But what if it is not?

X: What is the polish word for wreaths?

Y: t_4 - t_4

... a festival called Wianki (Polish for Wreaths) have become a tradition and a yearly event in the programme of cultural events in Warsaw. The festival traces its roots to a peaceful pagan ritual where maidens would float their wreaths of herbs on the water to predict when they would be married, and to whom ...

How to model this problem?

Let the output labels consist of **pointers to the input.**

$$\mathbf{x} = \{ x_1, \mathbf{x_2}, \mathbf{x_3}, x_4, x_5, \dots \mathbf{x_{n-1}}, x_n \}$$

$$\mathbf{y} = \{ 2, 3, n-1 \}$$

Which problems can be modeled this way?

- **Combinatorial Problems**
  Convex Hull; Traveling Salesman Problem
- **Annotation Based NLP**
  Named Entity Recognition; Argument Mining; Reading Comprehension …
- **What else?**

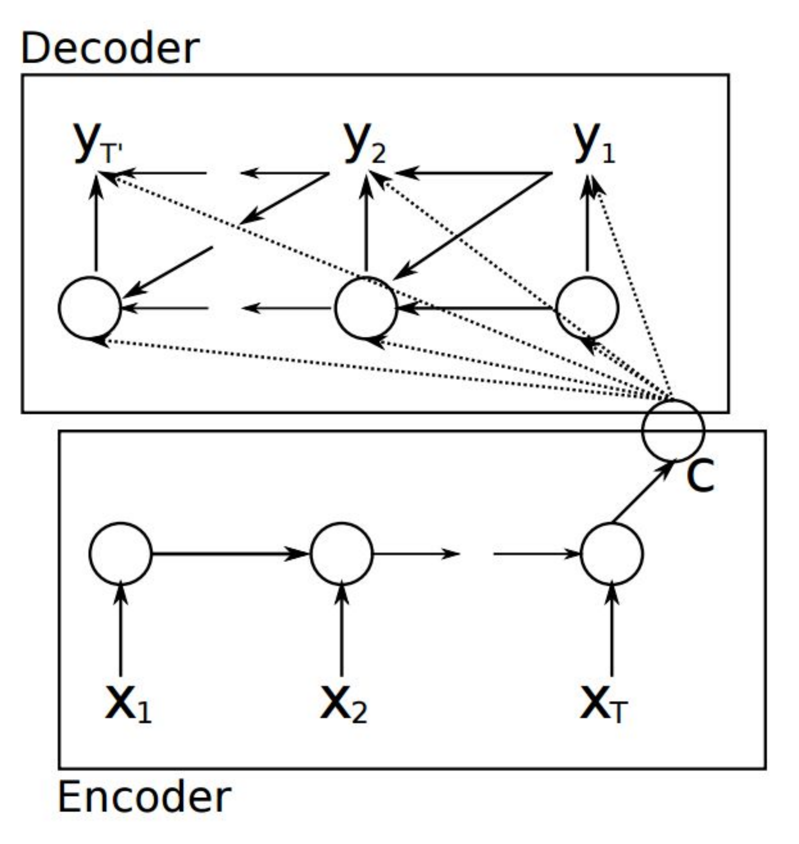
Attention Mechanism

A quick refresher.

Sequence to Sequence

A quick refresher.

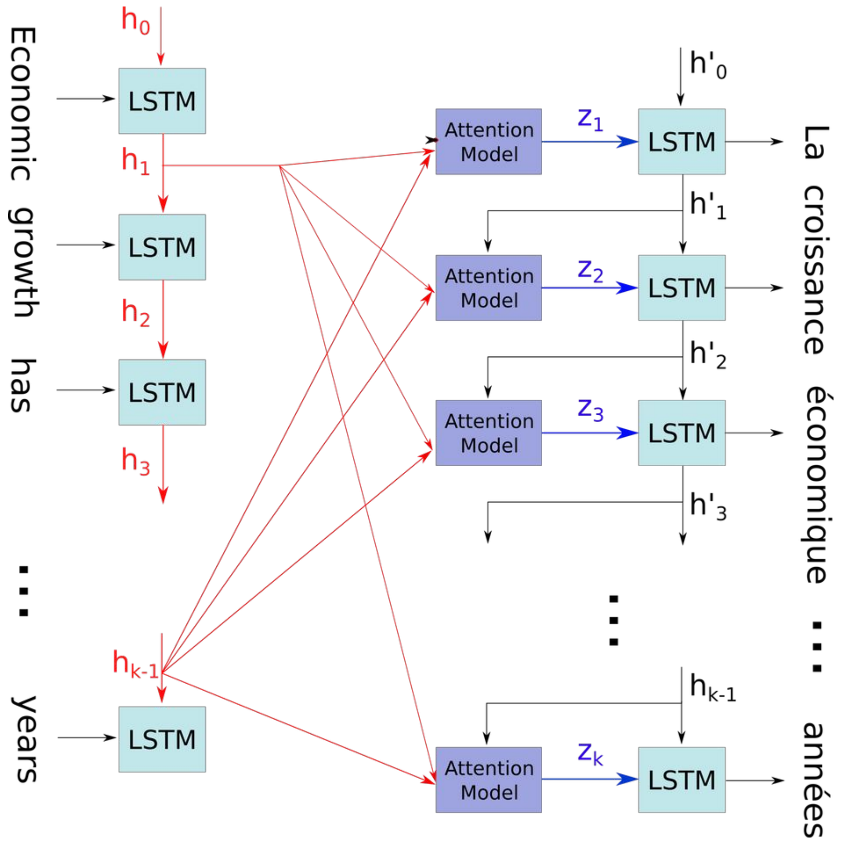
1. Encode the inputs via an encoder RNN.
2. After a certain point, the final hidden state of the encoder is used as a context vector for the decoder.
3. Using the context, decoder outputs desired labels.



Combine the encoder states to propagate extra information to the decoder.

Attention Mechanism

- ➔ Create a weighted combination of all input states.
  (Different weights for different position in output labels.)
- ➔ Major performance boost for data with longer term dependencies.



$$u_j^i = v^T tanh(W_1 e_j + W_2 d_i)$$

$$a_j^i = softmax(u_j^i)$$

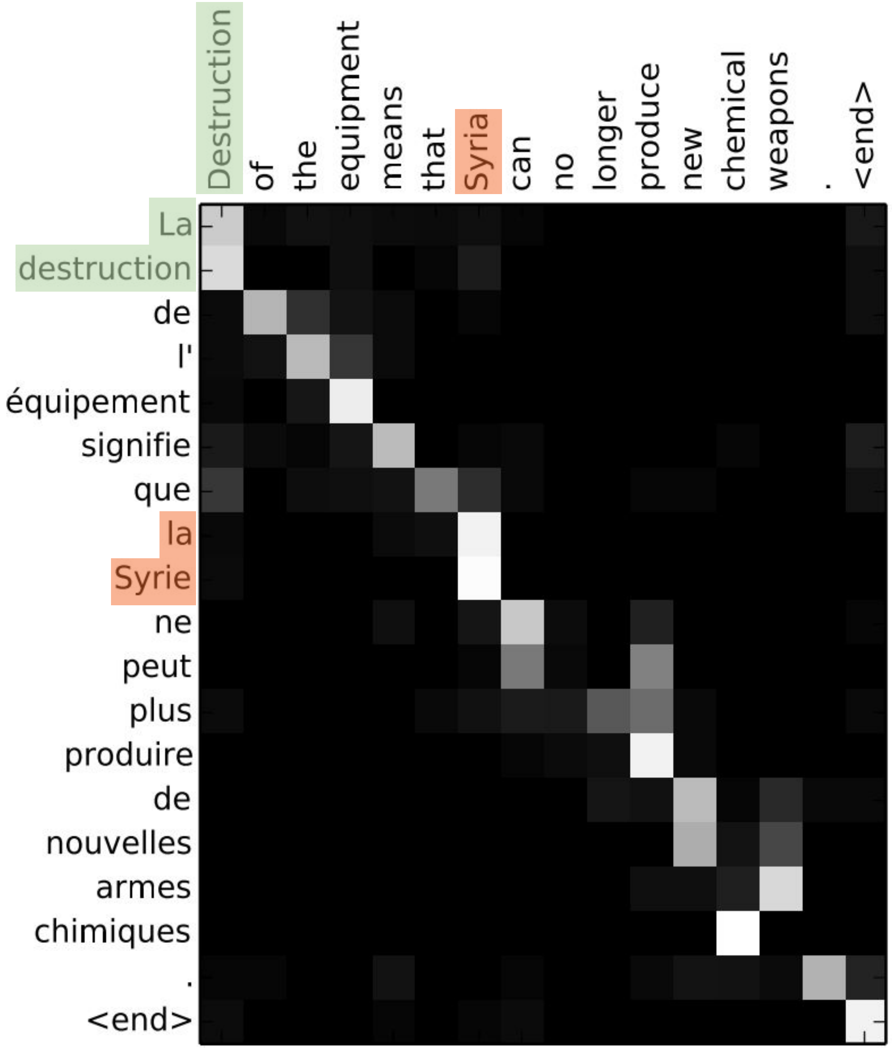
$$d_i' = \sum_{j=1}^{n} a_j^i e_j$$

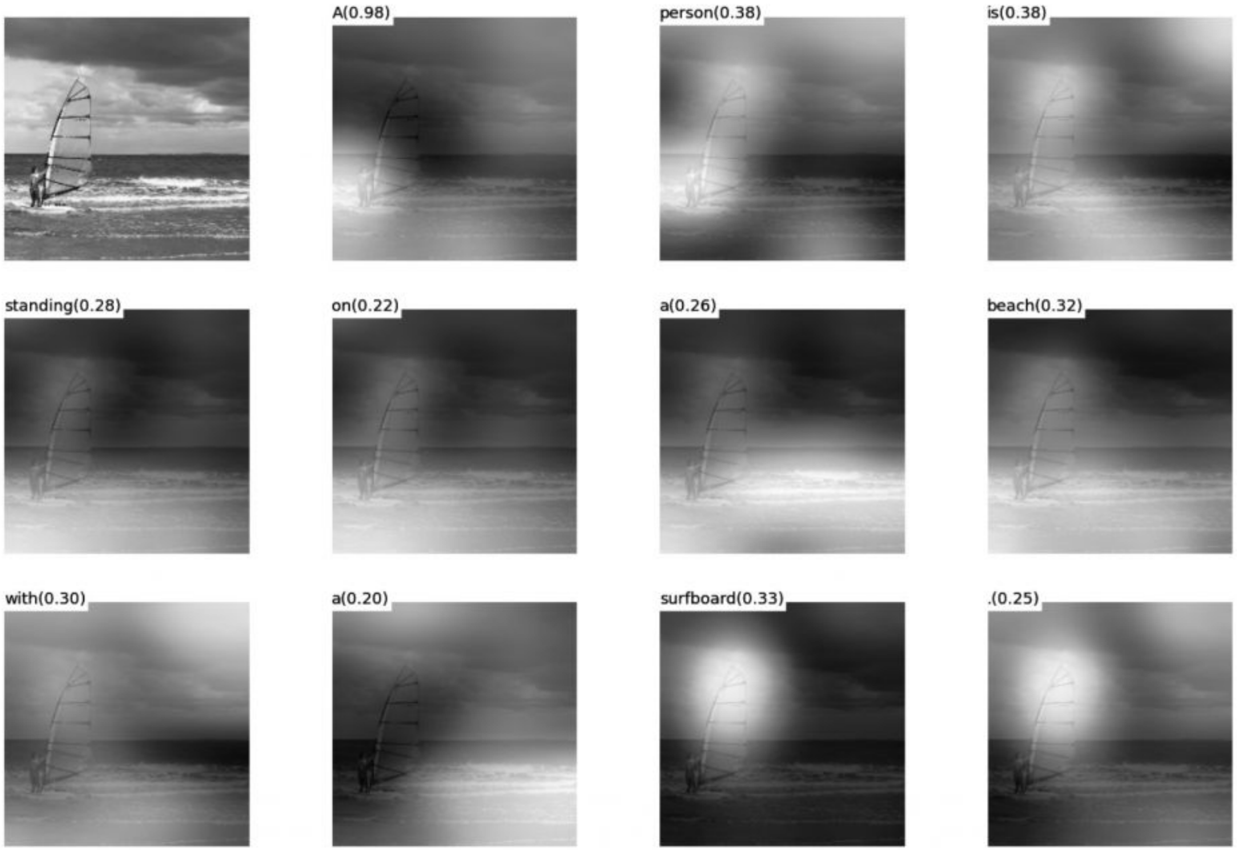
Visualized Attention

Attention can be visualized, for insights.

Interpreted as focus.

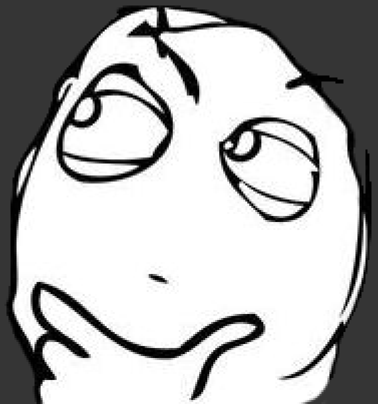
What part of the input is relevant for this part of the output.





(b) A person is standing on a beach with a surfboard.

… instead, use the attention vector as pointers to the input element …



Pointer Networks

Formally,

RNNs with Attention

$$u_j^i = v^T tanh(W_1 e_j + W_2 d_i)$$

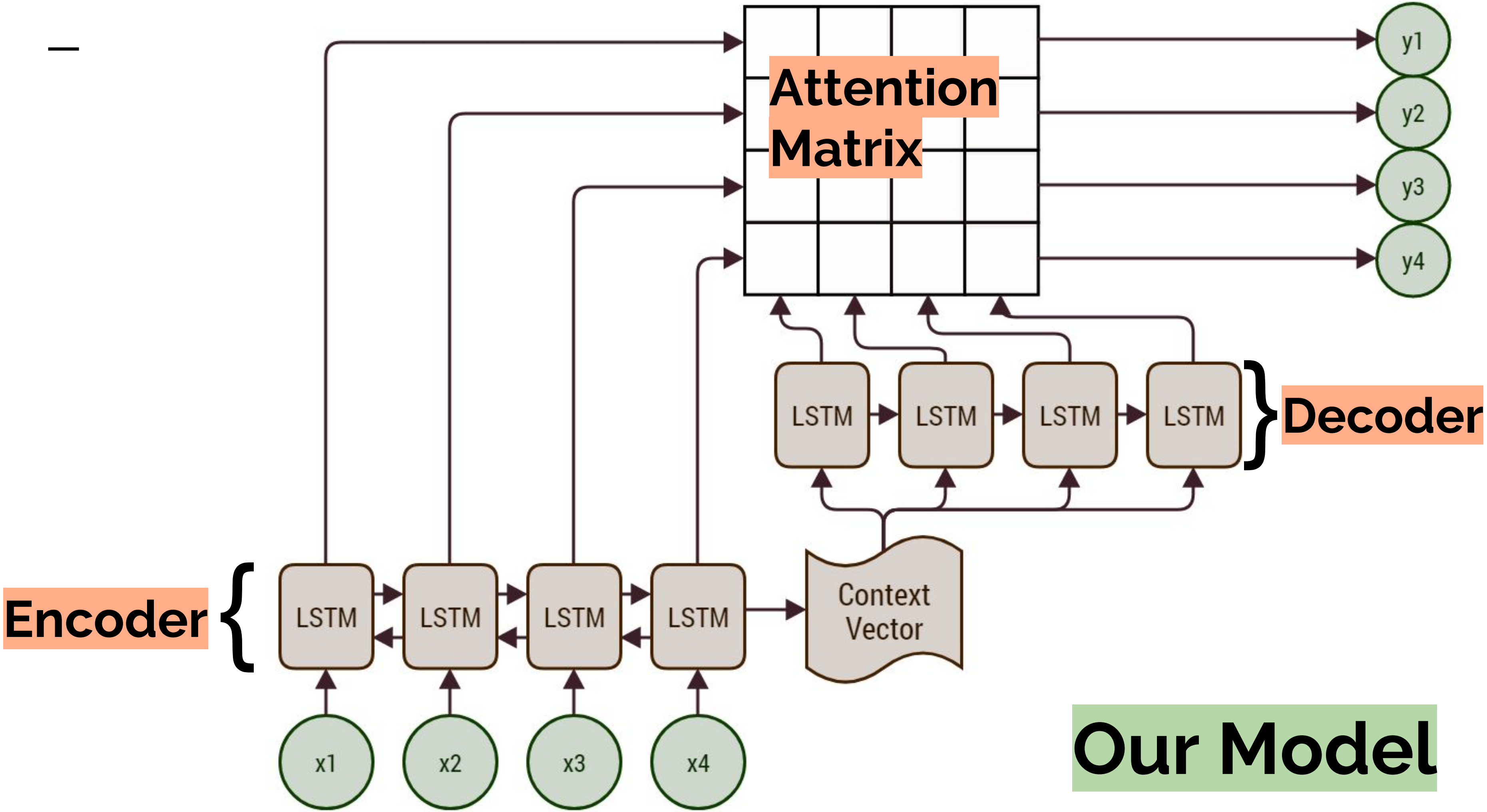
$$a_j^i = softmax(u_j^i)$$

$$\boxed{d_i'} = \sum_{j=1}^{n} a_j^i e_j$$

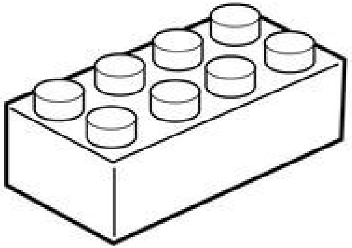
Pointer Networks

$$u_j^i = v^T tanh(W_1 e_j + W_2 d_i)$$

$$\boxed{a_j^i} = softmax(u_j^i)$$


Attention Matrix
y1
y2
y3
y4
LSTM
LSTM
LSTM
LSTM
Decoder
Encoder
LSTM
LSTM
LSTM
LSTM
Context Vector
x1
x2
x3
x4
Our Model

Experiments



Toy Problem

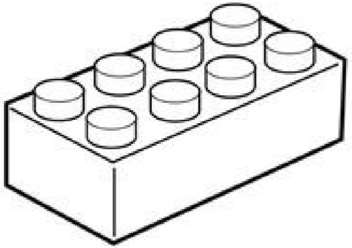
Input: sequence of integers < 5;

+ sequence of integers > 5, 10;

+ sequence of integers < 5

Variable Length of all three sequences.

Output: Find the middle sequence.



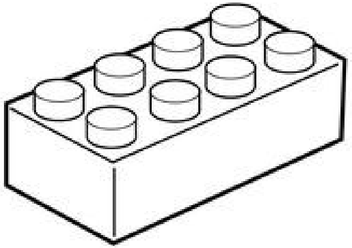
Toy Problem

x = { 4,1,2,3,1,1,6,9,10,8,7,3,1,1 }

y = { /start, /end }

x = { 1,2,2,1,5,9,7,1 }

y = { /start, /end }



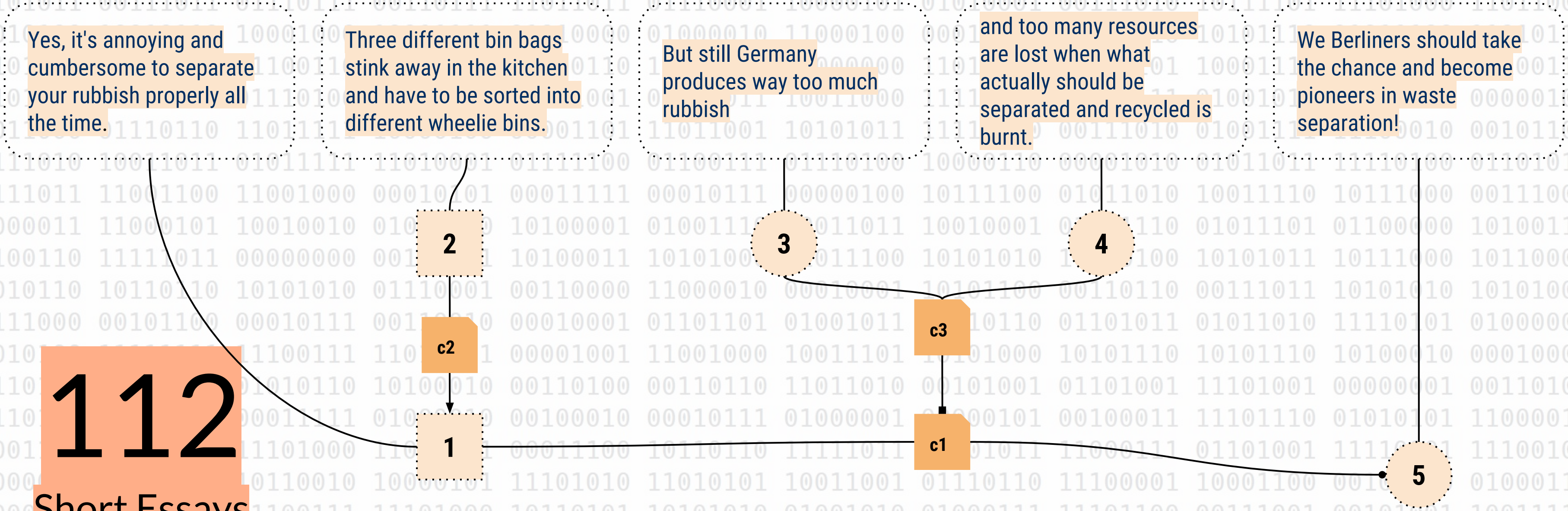
Toy Problem

97%

accuracy

Argumentation Mining: Dataset

Peldszus and Stede (2016)



112

Short Essays

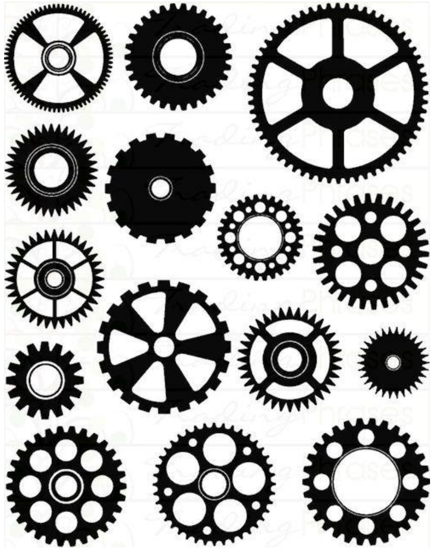
Argumentation Mining: Modeling Details

➔ Encode every word of the argument clause (AC) to word vectors.

➔ Compress these word vectors into a **Sentence Vector.**

➔ **X**: all the argument clauses (sentence vector for each AC) in different time steps.

➔ **Y**: an adjacency matrix representing the AC Links.



Argumentation Mining: Further Details

2,000
Epochs

512
Hidden Dimensions

300
Input Dimensions

0.001
Learning Rate

4h 15m
Training Time

5
Fold Validation

Argumentation Mining: Results

System	F1 Score
(Christian, and Gurevych, 2017)	.683
(Potash et. al, 2016)	.740
Our Model	**.750**

Argumentation Mining: Results

Models	F1 Score
Word2Vec: 300 dim, 2000 epochs	.750
GloVe: 300 dim, 2000 epochs	0.72

Our Contributions

1. We implemented Pointer Networks using Keras backend, with LSTM encoders and decoders.
    a. Evidently this is the only publicly available implementation on Keras.
    b. Our implementation can work both Theano and Tensorflow in the backend.
2. Test the Pointer Network implementation with a toy problem mentioned in this blog.
3. Augment the network of (Potash et. al, 2016) to achieve state of the art on this task.
4. Compare pre-trained embeddings (GloVe, Word2Vec) on our model.

Our Contributions

- Our repository: https://github.com/geraltofrivia/argument-mining
- Pointer Network implementation: https://github.com/saist1993/seq2seq
- Argument Mining reading list: https://goo.gl/LbY2cJ

Future Work

- ➔ Jointly predict AC classes as well as links.
- ➔ Explore different ways of embedding sentences.
- ➔ Jointly perform more than one task, by extending our network to see if it boosts the performance.
- ➔ Try to train the model on Persuasive Essays corpus, as well.
- ➔ Pre-train the model to mimic constraints on links.

Questions?


FISHING FOR ATTENTION
WE ARE
memegenerator.net

References

- ➔ Vinyals, Oriol, Meire Fortunato, and Navdeep Jaitly. **"Pointer networks."** Advances in Neural Information Processing Systems. 2015.
- ➔ Peldszus, Andreas, Manfred Stede. **“An annotated corpus of argumentative microtexts”.** First European Conference on Argumentation: Argumentation and Reasoned Action, Portugal, Lisbon, June 2015.
- ➔ Potash, Peter, Alexey Romanov, and Anna Rumshisky. "**Here's My Point: Argumentation Mining with Pointer Networks.**" arXiv preprint arXiv:1612.08994 (2016).
- ➔ Koreeda, Yuta, et al. "**Neural Attention Model for Classification of Sentences that Support Promoting/Suppressing Relationship.**" ACL 2016 (2016): 76.
- ➔ Raquel Mochales Palau and Marie-Francine Moens. 2011. **“Argumentation mining”.** Artificial Intelligence and Law, 19(1):1–22.
- ➔ Teufel, Simone, and Min-Yen Kan. "**Robust argumentative zoning for sensemaking in scholarly documents.**" Advanced Language Technologies for Digital Libraries. Springer, Berlin, Heidelberg, 2011. 154-170.
- ➔ Stab, Christian, and Iryna Gurevych. **"Parsing argumentation structures in persuasive essays."** Computational Linguistics (2017).
- ➔ Persing, I., & Ng, V. (2016). “**End-to-End Argumentation Mining in Student Essays**”. Naacl, 1384–1394. Retrieved from http://www.aclweb.org/anthology/N16-1164

- ➔ Further Reading: [This blog post] (Pointer Networks: authored by Dev Nag)