

Building a QA System using Deep Learning Techniques

Rricha Jalota
DICE Research Group, Paderborn University

<https://www.limbo-project.org/>

Linked Data

Legend

Cross Domain

Geography

Government

Life Sciences

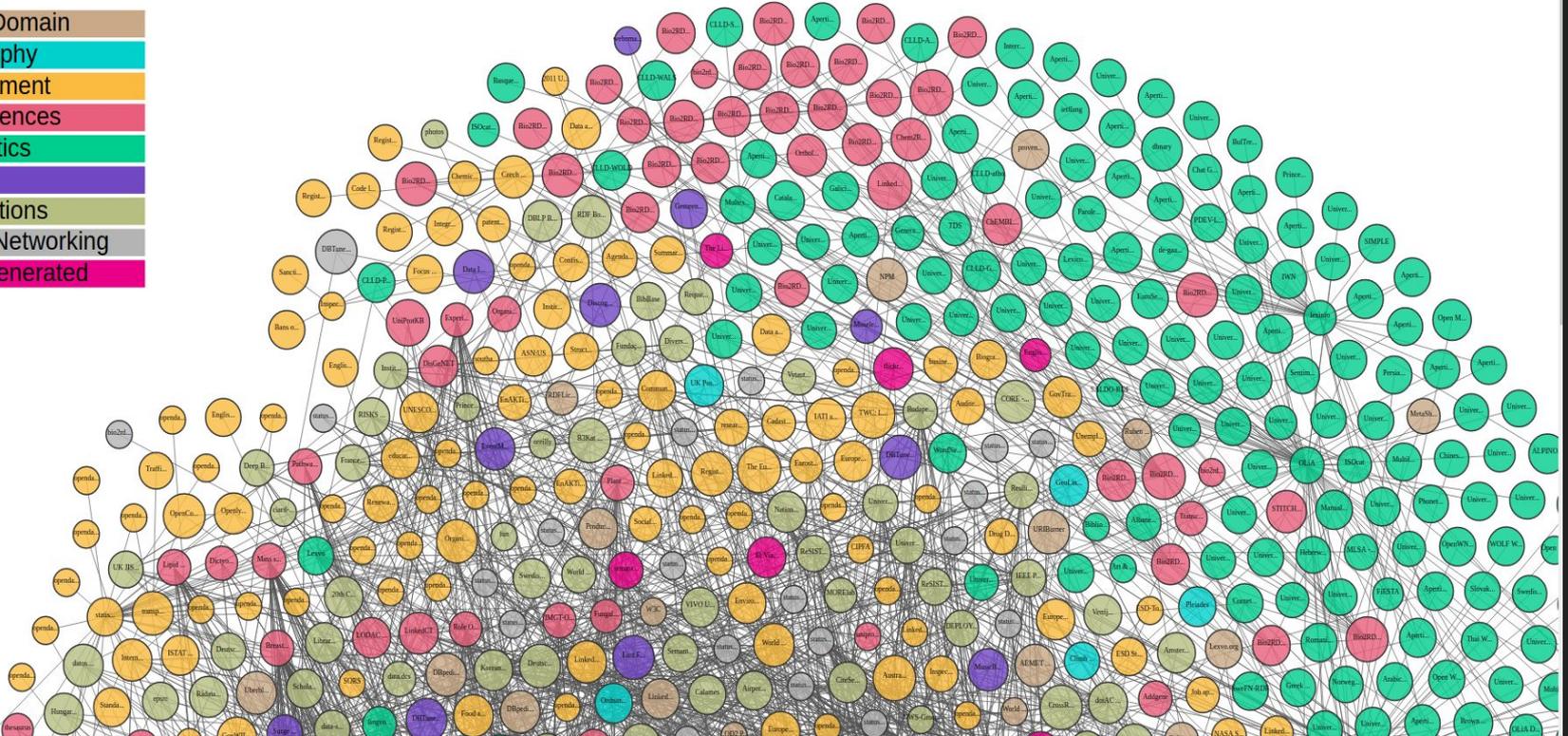
Linguistics

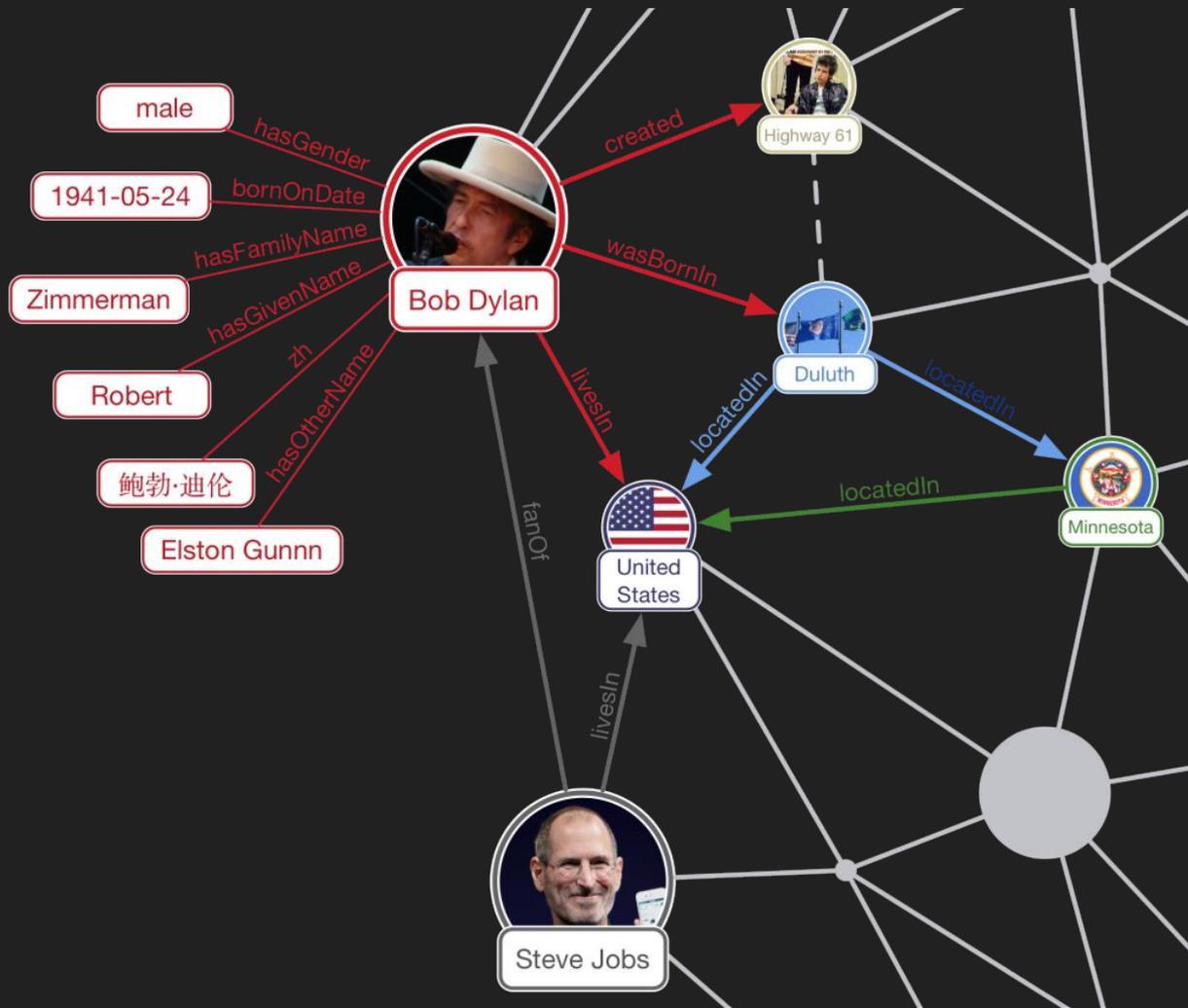
Media

Publications

Social Networking

User Generated





<https://github.com/dice-group>

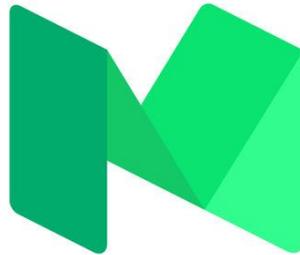
Unstructured Data

The Three Little Pigs

From Wikipedia, the free encyclopedia
(Redirected from [Three little pigs](#))

For other uses, see [Three Little Pigs \(disambiguation\)](#).

The Three Little Pigs is a *fable* about three pigs who build three houses of different materials. A **big bad wolf** blows down the first two pigs' houses, made of **straw** and **sticks** respectively, but is unable to destroy the third pig's house, made of **bricks**. Printed versions date back to the 1840s, but the story itself is thought to be much older. The phrases used in the story, and the various **morals** drawn from it, have become embedded in **Western culture**. Many versions of *The Three Little Pigs* have been recreated or have been modified over the years, sometimes making the wolf a kind character. It is a type 124 folktale in the **Aarne–Thompson classification system**.

The Weebly logo is displayed in white, lowercase, rounded letters with a slight shadow effect, set against a solid blue background.

WORDPRESS

Reading Comprehension



Name _____

Read the following story and answer the questions below.

Jack is a seven year old boy who loves adventure. He likes to climb trees and find interesting bugs and leaves in his backyard. He also likes to play with his dog, Daisy. Daisy is a two year old yellow lab mix. Jack and Daisy are best friends.

1. How old is Jack?

He is _____ .

2. What does Jack like to do in his backyard?

He likes to _____ .

3. What kind of a dog is Daisy?

She is a _____ .



How is it done?

Why is it done?

Challenges?

Roadmap

1. Familiarize ourselves with available datasets
 - a. Explore the selected dataset
2. Discuss intuitive NLP approaches
3. Dive into Deep Learning
 - a. Some theory
 - b. Approach Formulation
 - c. Some code!
4. Research

<https://github.com/dice-group/EuroPython-2018>

Available Datasets

- DeepMind QA Dataset : CNN/Daily Mail RC corpus
 - 1.5 million question and answer pairs
 - Cloze dataset
 - Goal: Predict the missing word (named entity) in the passage

Passage

(@entity4) if you feel a ripple in the force today , it may be the news that the official @entity6 is getting its first gay character . according to the sci-fi website @entity9 , the upcoming novel " @entity11 " will feature a capable but flawed @entity13 official named @entity14 who " also happens to be a lesbian . " the character is the first gay figure in the official @entity6 -- the movies , television shows , comics and books approved by @entity6 franchise owner @entity22 -- according to @entity24 , editor of " @entity6 " books at @entity28 imprint @entity26 .

Question

characters in " @placeholder " movies have gradually become more diverse

Answer

@entity6

Model	CNN		Daily Mail	
	Dev	Test	Dev	Test
Frame-semantic model [†]	36.3	40.2	35.5	35.5
Word distance model [†]	50.5	50.9	56.4	55.5
Deep LSTM Reader [†]	55.0	57.0	63.3	62.2
Attentive Reader [†]	61.6	63.0	70.5	69.0
Impatient Reader [†]	61.8	63.8	69.0	68.0
MemNNs (window memory) [‡]	58.0	60.6	N/A	N/A
MemNNs (window memory + self-sup.) [‡]	63.4	66.8	N/A	N/A
MemNNs (ensemble) [‡]	66.2*	69.4*	N/A	N/A
Ours: Classifier	67.1	67.9	69.1	68.3
Ours: Neural net	72.5	72.7	76.9	76.0
Ours: Neural net (ensemble)	76.2*	76.5*	79.5*	78.7*
Ours: Neural net (relabeling)	73.8	73.6	77.6	76.6
Ours: Neural net (relabeling, ensemble)	77.2*	77.6*	80.2*	79.2*

Table 2: Accuracy of all models on the *CNN* and *Daily Mail* datasets. Results marked [†] are from (Hermann et al., 2015) and results marked [‡] are from (Hill et al., 2016). *Classifier* and *Neural net* denote our entity-centric classifier and neural network systems respectively. The numbers marked with * indicate that the results are from ensemble models.

- Children's Book Test (CBT)
 - 677,343 million context-query pairs
 - Context: first 20 sentences
 - Query: A word removed from the 21st sentence
 - Goal: Identify the answer word from 10 candidate answers appearing in the context sentences and the query

- Other Datasets: MCTest, MS Macro, etc

```

Context:
1 So they had to fall a long way .
2 So they got their tails fast in their mouths .
3 So they could n't get them out again .
4 That 's all .
5 `` Thank you , " said Alice , `` it 's very interesting .
6 I never knew so much about a whiting before . "
7 `` I can tell you more than that , if you like , " said the Gryphon .
8 `` Do you know why it 's called a whiting ? "
9 `` I never thought about it , " said Alice .
10 `` Why ? "
11 `` IT DOES THE BOOTS AND SHOES . '
12 the Gryphon replied very solemnly .
13 Alice was thoroughly puzzled .
14 `` Does the boots and shoes ! "
15 she repeated in a wondering tone .
16 `` Why , what are YOUR shoes done with ? "
17 said the Gryphon .
18 `` I mean , what makes them so shiny ? "
19 Alice looked down at them , and considered a little before she gave her answer .
20 `` They 're done with blacking , I believe . "

Query: `` Boots and shoes under the sea , " the XXXXX went on in a deep voice , ``

Candidates: Alice|BOOTS|Gryphon|SHOES|answer|fall|mouths|tone|way|whiting

Answer: gryphon

```

- SQuAD-v1.1 / v2.0
 - 100,000 questions on Wikipedia articles
 - 50,000 questions added
 - Determine when no answer is supported

Passage Segment

...The European Parliament and the Council of the European Union have powers of amendment and veto during the legislative process...

Question

Which governing bodies have veto power?

Question Answering, SQuAD v1.1



Source: stanford-qa.com

SQuAD1.1 Leaderboard

Since the release of SQuAD1.0, the community has made rapid progress, with the best models now rivaling human performance on the task. Here are the ExactMatch (EM) and F1 scores evaluated on the test set of v1.1.

Rank	Model	EM	F1
	Human Performance <i>Stanford University</i> (Rajpurkar et al. '16)	82.304	91.221
1 Jul 12, 2018	QANet (ensemble) <i>Google Brain & CMU</i>	84.454	90.490
2 Jul 09, 2018	r-net (ensemble) <i>Microsoft Research Asia</i>	84.003	90.147
3 Jun 21, 2018	MARS (ensemble) <i>YUANFUDAO research NLP</i>	83.982	89.796
4 Mar 20, 2018	QANet (ensemble) <i>Google Brain & CMU</i>	83.877	89.737
5 Jun 21, 2018	MARS (single model) <i>YUANFUDAO research NLP</i>	83.122	89.224
6 Mar 07, 2018	QANet (ensemble) <i>Google Brain & CMU</i>	82.744	89.045
7 May 09, 2018	MARS (single model) <i>YUANFUDAO research NLP</i>	82.587	88.880

Leaderboard

SQuAD2.0 tests the ability of a system to not only answer reading comprehension questions, but also abstain when presented with a question that cannot be answered based on the provided paragraph. How will your system compare to humans on this task?

Rank	Model	EM	F1
	Human Performance <i>Stanford University</i> (Rajpurkar & Jia et al. '18)	86.831	89.452
1 Jul 13, 2018	VS^3-NET (single model) <i>Kangwon National University in South Korea</i>	68.438	71.282
2 Jun 25, 2018	KACTEIL-MRC(GFN-Net) (single model) <i>Kangwon National University, Natural Language Processing Lab.</i>	68.224	70.871
3 Jun 26, 2018	KakaoNet2 (single model) <i>Kakao NLP Team</i>	65.708	69.369
4 Jul 11, 2018	abcNet (single model) <i>Fudan University & Liulishuo AI Lab</i>	65.256	69.198
5 Jun 27, 2018	BSAE AddText (single model) <i>reciTAL.ai</i>	63.383	67.478

Let's explore, SQuAD!

Baseline Approach

Given: context, question, answer and answer's start index/answer span

1. Generate answer candidates: Only use spans which are constituents in the constituency parse
2. Use sliding window based approach - keep all candidates that have the maximal (unigram/bigram) overlap - select the best one!
3. Logistic Regression

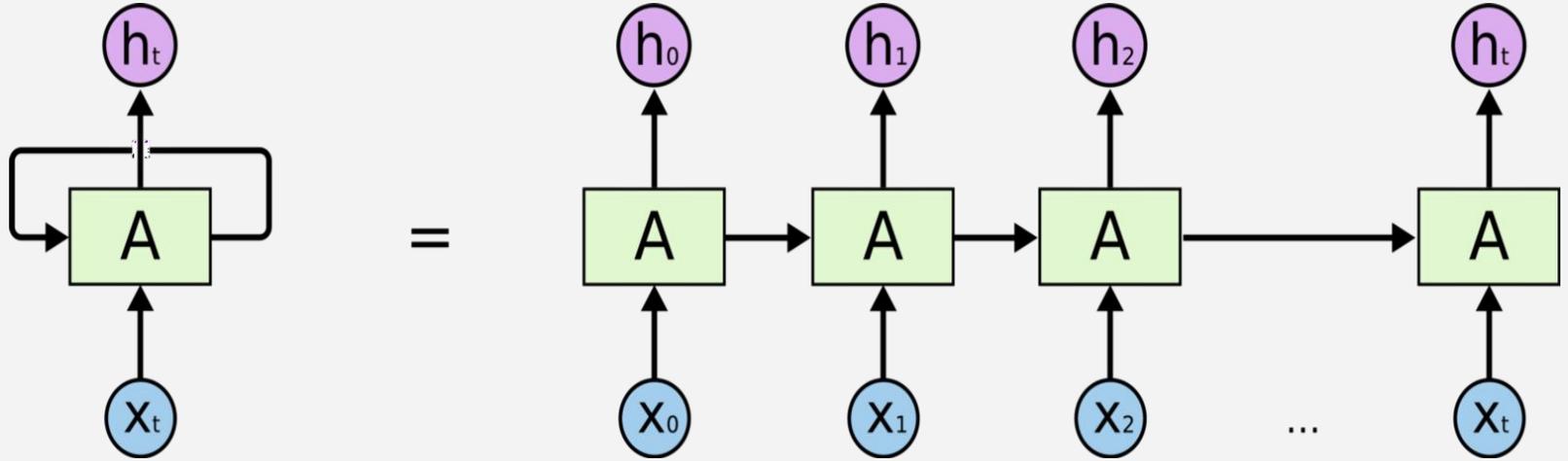
Dive into Deep Learning

1. Preprocess the Dataset

- a. Convert text input into a numeric representation
- b. Word Representation can be
 - i. One-hot Encoding - sparse representation
 - ii. Word Embeddings - dense representation of words and their relative meanings
 1. Map semantic meaning into a geometric space
 2. Geometric space - Embedding space
 3. Word2Vec, gloVe

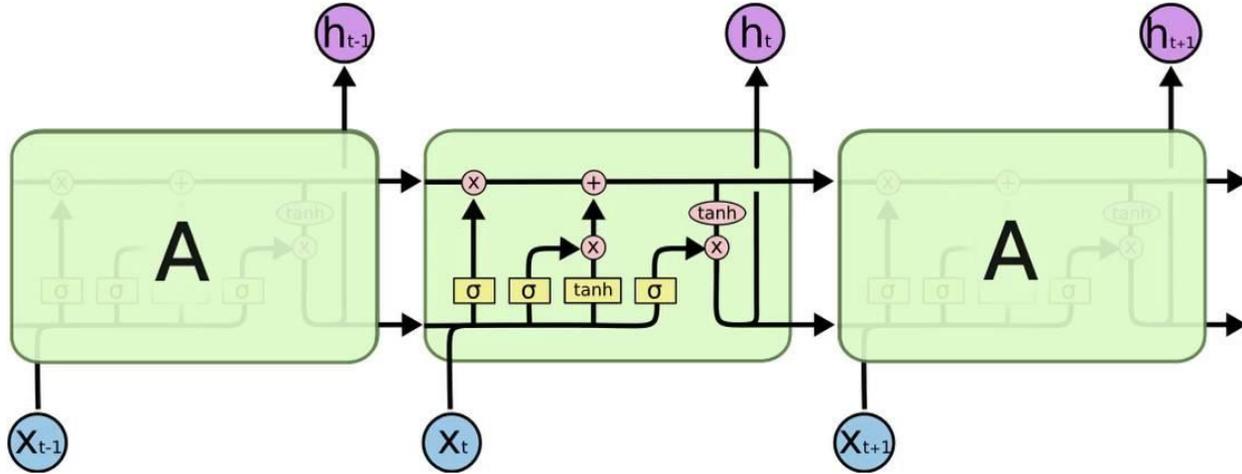
2. Encoder - Decoder

RNNs

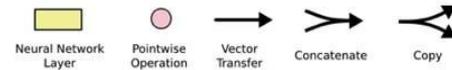


LSTMs

Long-Short Term Memory module: LSTM



long-short term memory modules used in an RNN



Let's look into the code!

Thank You for your Attention!

Get in touch: rricha.jalota@uni-paderborn.de
<https://twitter.com/RichaJalota> <https://twitter.com/DiceResearch>

Feedback form : <https://goo.gl/forms/4Cg4p3AqVIP17Lon2>

Github repo : <https://github.com/dice-group/EuroPython-2018>

Interested in Linked Data Services for Mobility? Check out
<https://www.limbo-project.org/>

