Learning Transferable Features for Open-Domain Question Answering

Gianlucca Zuin Luiz Chaimowicz Adriano Veloso

Universidade Federal de Minhas Gerias
Departamento de Ciência da Computação

July 11, 2018

IEEE WORLD CONGRESS ON COMPUTATIONAL INTELLIGENCE
WCCI 2018
Rio de Janeiro, Brazil | July 8-13, 2018

DCC
DEPARTAMENTO DE CIÊNCIA DA COMPUTAÇÃO

UFMG

Overview

Introduction

Problem Definition

A question answering system is a system capable of receiving as input a question in natural language and that attempts to return an answer also in natural language.

Problem Definition

We consider the task of learning open-domain Question Answering models (hereinafter QA models), that is, QA models that find answers in collections of unstructured documents to questions about nearly anything or any topic.

Motivation

- Difficult to collect large-scale corpora for open-domain QA models.
- Restricting the question domain, the data demand becomes significantly smaller.

Key insight: an open-domain QA model can be decomposed into multiple domain-specific QA models, each being learnt independently.

Our approach

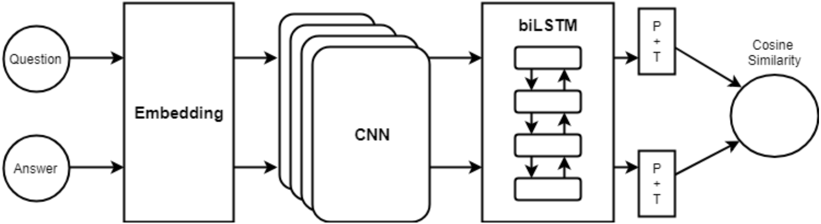
- Individual questions can be mapped to specific topic domains.
- A topic domain is defined in terms of common words shared by all domains and words that are specific to individual domains.
- QA models are learned for each domain.

However learning domain-specific QA models is still challenging...

Solution: *Domain Adaptation*.

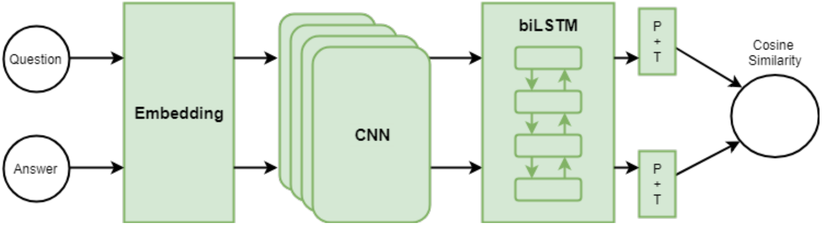
Neural Network

Network Architecture

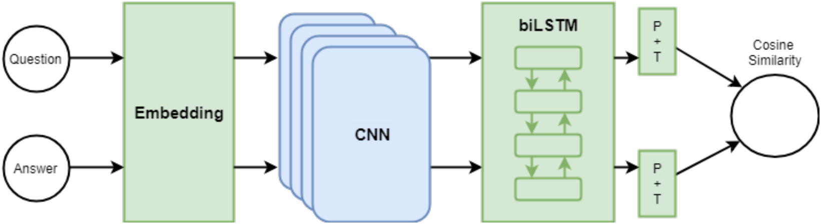


Complementary semantic perspective: temporal information of textual features in question-answer pairs.

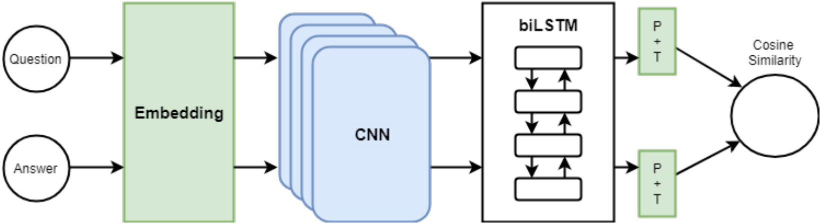
Transfer Learning: T1



Transfer Learning: T2



Transfer Learning: T3



Dataset

SQuAD Dataset

"Stanford Question Answering Dataset (SQuAD) is a new reading comprehension dataset, consisting of questions posed by crowdworkers on a set of Wikipedia articles, where the answer to every question is a segment of text, or span, from the corresponding reading passage. With 100,000+ question-answer pairs on 500+ articles, SQuAD is significantly larger than previous reading comprehension datasets."

Span-level QA

Amazon_rainforest

The Stanford Question Answering Dataset

How many square kilometers of rainforest is covered in the basin?

Ground Truth Answers: 5,500,000 square kilometres (2,100,000 sq mi)

The Amazon rainforest (Portuguese: Floresta Amazônica or Amazônia; Spanish: Selva Amazónica, Amazonía or usually Amazonia; French: Forêt amazonienne; Dutch: Amazoneregenwoud), also known in English as Amazonia or the Amazon Jungle, is a moist broadleaf forest that covers most of the Amazon basin of South America. This basin encompasses 7,000,000 square kilometres (2,700,000 sq mi), of which 5,500,000 square kilometres (2,100,000 sq mi) are covered by the rainforest. This region includes territory belonging to nine nations. The majority of the forest is contained within Brazil, with 60% of the rainforest, followed by Peru with 13%, Colombia with 10%, and with minor amounts in Venezuela, Ecuador, Bolivia, Guyana, Suriname and French Guiana. States or departments in

Sentence-level QA

Amazon_rainforest

The Stanford Question Answering Dataset

How many square kilometers of rainforest is covered in the basin?

Ground Truth Answers: 5,500,000 square kilometres (2,100,000 sq mi)

The Amazon rainforest (Portuguese: Floresta Amazônica or Amazônia; Spanish: Selva Amazónica, Amazonía or usually Amazonia; French: Forêt amazonienne; Dutch: Amazoneregenwoud), also known in English as Amazonia or the Amazon Jungle, is a moist broadleaf forest that covers most of the Amazon basin of South America. This basin encompasses 7,000,000 square kilometres (2,700,000 sq mi), of which 5,500,000 square kilometres (2,100,000 sq mi) are covered by the rainforest. This region includes territory belonging to nine nations. The majority of the forest is contained within Brazil, with 60% of the rainforest, followed by Peru with 13%, Colombia with 10%, and with minor amounts in Venezuela, Ecuador, Bolivia, Guyana, Suriname and French Guiana. States or departments in

Domain division

	Train		Test	
Domains	Articles	Questions	Articles	Questions
Person	48	12399	5	4697
History	38	10060	5	3607
City	29	7784	3	1478
Entertainment	33	6479	2	1284
Biology	37	6351	4	2480
Location	36	6139	5	3334
Technology	32	5094	4	2814
Law	27	3840	5	3592
Religion	18	3788	2	1168
Sports	16	3478	2	3172
Organization	17	2854	1	648
Thing	16	2845	1	249
Education	16	2527	5	2053
State	11	2527	2	1062
Science	13	2128	5	4356

Experiments

Experiments

RQ1: Does domain adaptation improve the effectiveness of our CNN-biLSTM models for span-level QA?

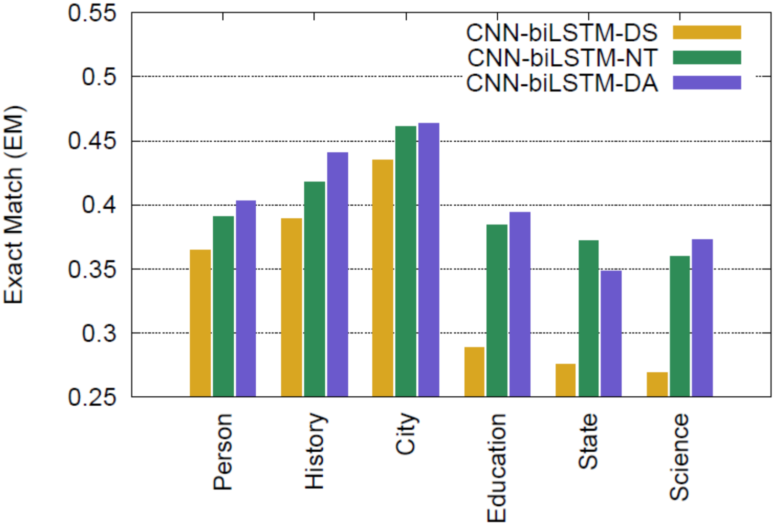
RQ2: Which feature transference approach is more appropriate to each topic domain?

RQ3: Does sentence-level information improve span-level QA performance?

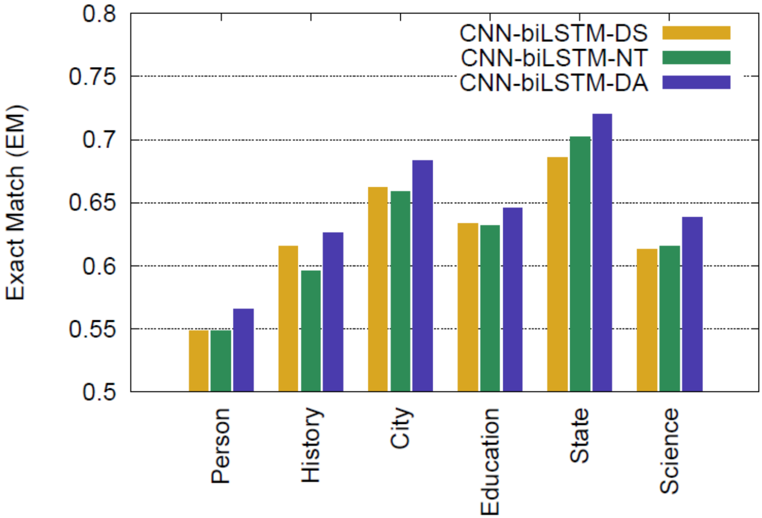
RQ4: What is the impact of applying simple topic identification methods?

RQ5: How do our CNN-biLSTM models compare against existing models?

Experiments



Experiments



Experiments

RQ1: Does domain adaptation improve the effectiveness of our CNN-biLSTM models for span-level QA?

RQ2: Which feature transference approach is more appropriate to each topic domain?

RQ3: Does sentence-level information improve span-level QA performance?

RQ4: What is the impact of applying simple topic identification methods?

RQ5: How do our CNN-biLSTM models compare against existing models?

Experiments

Domain	f^d_{span} T1	f^d_{span} T2	f^d_{span} T3	f^d_{sent} T1	f^d_{sent} T2	f^d_{sent} T3	best combination $f^d_{span} + f^d_{sent}$
Person	**.405**	**.404**	**.402**	**.561**	**.565**	**.561**	.504
History	.439	.449	**.456**	**.625**	**.624**	.619	.590
City	.459	**.469**	.463	**.683**	.672	.670	.656
Entertainment	**.381**	.361	.373	.578	**.584**	.572	.519
Biology	**.296**	.274	.279	**.591**	.574	.576	.444
Location	.377	**.391**	**.387**	.644	**.648**	.640	.513
Technology	.336	**.344**	.334	.630	.627	**.641**	.483
Law	.320	**.323**	.317	.604	.603	**.611**	.470
Religion	**.365**	**.368**	**.364**	**.691**	**.688**	.611	.542
Sports	**.389**	.388	**.392**	.633	.639	**.702**	.492
Organization	**.554**	.528	.522	.628	.642	**.662**	.671
Thing	.346	**.374**	.329	**.772**	**.775**	.768	.609
Education	.378	.394	**.399**	**.645**	.638	**.646**	.522
State	**.398**	**.400**	.388	.707	**.720**	.714	.597
Science	**.372**	**.373**	.369	.625	.621	**.638**	.503
Minimum	.296	.274	.279	.561	.565	.561	.444
Maximum	.554	.528	.522	.772	.775	.768	.671

Experiments

RQ1: Does domain adaptation improve the effectiveness of our CNN-biLSTM models for span-level QA?

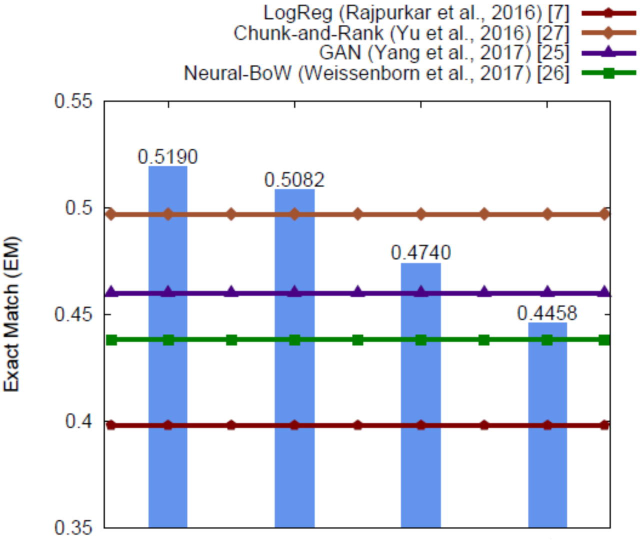
RQ2: Which feature transference approach is more appropriate to each topic domain?

RQ3: Does sentence-level information improve span-level QA performance?

RQ4: What is the impact of applying simple topic identification methods?

RQ5: How do our CNN-biLSTM models compare against existing models?

Experiments



Conclusions

Conclusions

- No feature transference approach is superior than the others.
- Domain adaptation using implicit topic information leads to improved QA performance
- A simple approach to condition the choice of the answer based on sentences that are relevant to the question can be effective.

Conclusions

Our results indicate that domain adaptation is highly effective, leading to accuracy gains that are as high as 20% in some domains. On average, our models have a 10% increase in accuracy by performing domain adaptation. Sentence conditioning is also very effective, as we observed a 40% increase in the span-level QA performance when using sentence relevance information while selecting the answer.

Learning Transferable Features for Open-Domain Question Answering

Gianlucca Zuin Luiz Chaimowicz Adriano Veloso

Universidade Federal de Minhas Gerias
Departamento de Ciência da Computação

July 11, 2018

IEEE WORLD CONGRESS ON COMPUTATIONAL INTELLIGENCE
WCCI 2018
Rio de Janeiro, Brazil | July 8-13, 2018

DCC
DEPARTAMENTO DE CIÊNCIA DA COMPUTAÇÃO

UFMG