Comparison of Transfer-Learning Approaches for Response Selection in Multi-Turn Conversations

Jesse Vig and Kalai Ramea

DSTC 7 Workshop, AAAI ‘19
Honolulu, HI
Jan. 27, 2019
Recent Trends in Transfer Learning for NLP
Task 1 Model

Language Representation

→

Task 2 Model

Language Representation
Word Embeddings (2013)

Word2Vec / GloVe / FastText

[I] [pet] [the] dog [and] [it] [barked]

Sentiment

<table>
<thead>
<tr>
<th></th>
<th>92%</th>
<th>3%</th>
<th>5%</th>
</tr>
</thead>
<tbody>
<tr>
<td>negative</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>neutral</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>positive</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Task Model

This film really dragged
Contextual Embeddings (2017)

**Word:** ELMo, CoVe

**(Sentence:** USE, GPDS, InferSent)

**ELMo**

---

**Task Model**

- 92% negative
- 3% neutral
- 5% positive

---

From "NLP's ImageNet Moment Has Arrived"
This film really dragged

Task head

| 92% negative | 3% neutral | 5% positive |

Layer 1

Layer N-1

Layer N

Fine-tuning (2018)

ULMFiT, OpenAI GPT, BERT, LM-LSTM (2015)
Comparison of Transfer Learning Approaches for DSTC7, Track 1
Student: I’m looking for an engineering course.
Advisor: How about CS481?
Student: I already have a heavy course load.
Advisor: I suggest CS425.

Candidate responses:
- I suggest CS425
- CS221 is not too demanding
- Nice talking to you
- What are you interested in?

Ranked responses:

<table>
<thead>
<tr>
<th>Rank</th>
<th>Response</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>CS221 is not too demanding</td>
</tr>
<tr>
<td>2</td>
<td>I suggest CS425</td>
</tr>
<tr>
<td>...</td>
<td>...</td>
</tr>
<tr>
<td>100</td>
<td>Hi, how can I help you?</td>
</tr>
</tbody>
</table>
Student: I’m looking for an engineering course
Advisor: How about CS481?
Student: I already have a heavy course load
Advisor: 

Candidate response

I suggest CS425
Hi, how can I help you?
CS221 is not too demanding
Nice talking to you
What are you interested in?

Correct
Incorrect

0.87 0.13
Models Evaluated

Submitted Model
Multi-turn ESIM+ELMo (MT-EE)

Feature-based

Multi-Turn ESIM*

Pre-trained, Fine-tuned

OpenAI GPT**

Enhanced Sequential Inference Model

MT-EE

BERT***

*Pre-trained, Fine-tuned

**Generative Pretrained Transformer

***Bidirectional Encoder Representations from Transformers
Multi-Turn ESIM+ELMo (MT-EE)

- Position-based weighted sum
- Classifier

- Speaker-specific feedforward
- Context utterances

ESIM

ELMo

utterance_0 response

utterance_{T-1} response

\[ \alpha_0 \]

\[ \alpha_{T-1} \]
OpenAI GPT

• Unidirectional self-attention
• Standard language model pretraining
• BooksCorpus
• Designed for arbitrary text inputs
  • Single sentence, two sentences, multiple choice, etc.
  • Accomplished through delimiter tokens

BERT

• Bidirectional self-attention
• “Masked” language model pretraining
• BooksCorpus + Wikipedia
• Optimizations for sentence pairs
  • Architecture
    • Segment embedding
  • Pre-training
    • Next sentence prediction
Student: Do I need to study any math courses?

Advisor: ?

BERT

Candidate response: You have completed all math required for your degree
do I need to study any math courses? you have completed all math required for your degree.
Sentence A

[CLS] do i need to study any math courses?

[SEP] you have completed all math courses?

Sentence B

[CLS] you have completed all math courses?

[SEP]

Sentence A

[CLS] do i need to study any math courses?

[SEP] you have completed all math courses?

Sentence B

[CLS] do i need to study any math courses?

[SEP] you have completed all math courses.

[SEP]
Resources

BERT visualization tool
https://github.com/jessevig/bertviz

*The Illustrated BERT, Elmo, and Co.*
http://jalammar.github.io/illustrated-bert/

*The Illustrated Transformer*
http://jalammar.github.io/illustrated-transformer/