MovieQA: Understanding Stories in Movies through Question-Answering

Answering and Evaluation

For any questions, email tapaswi@kit.edu or fidler@cs.toronto.edu

Registration and submissions are open!
Benchmark: http://movieqa.cs.toronto.edu

Generic QA framework
- all multiple-choice QA approaches

Data representations
- represent data in vector space
  - TF-IDF facilitates matching exact words
  - Word2Vec meanings of words, allows synonyms
  - SkipThought encodes semantics of sentence
- EncodeClip identifies objects/places, embeds in text space

SS with Convolutional Brain
- core idea
  - learn the three-way scoring function
  - weighted combination of scores from (story, question) and (story, answer)
- 1 × 1 convolutions

Evaluation
- plot-based answering easier, words repeated
- simple cosine similarity does not work with DVS, subtitles, scripts
- memory networks able to leverage this info.
- SSCB easily fuses all text representations

Method | Plot | DVS | Subtitle | Script
--- | --- | --- | --- | ---
Cosine TF-IDF | 47.6 | 24.5 | 24.5 | 24.6
Cosine Word2Vec | 46.4 | 26.6 | 24.5 | 23.4
Cosine SkipThought | 31.0 | 19.9 | 21.3 | 21.2
SSCB TF-IDF | 48.5 | 24.5 | 27.6 | 26.1
SSCB Word2Vec | 45.1 | 24.8 | 24.8 | 25.0
SSCB SkipThought | 28.3 | 24.5 | 20.8 | 21.0
SSCB Fusion | 56.7 | 24.8 | 27.7 | 28.7
MemN2N 1 layer | 40.6 | 33.0 | 38.0 | 42.3
MemN2N 3 layers | 42.3 | 33.0 | 37.1 | 43.0

The Hasty Student
- answer questions, don’t look at story
- pick correct answer as the
  - longest answer
  - most similar/distinct answer
  - answer most similar to the question

The Hasty Turker
- 10 AMT workers answer questions without looking at the story

The Searching Student
- search within the story to find the best match for question and the answer
- windowed cosine similarity to compute how well a story fits a Q and A

The Hasty Turker
- 10 AMT workers answer questions without looking at the story

For all multiple-choice QA approaches

use a three-way function between story, question, and answers
e.g. a CNN-RNN approach on VQA
1. CNN(story = image); 2. RNN(question); 3. answer = softmax(vocabulary)

MovieQA: Understanding Stories in Movies through Question-Answering

Answering and Evaluation

For any questions, email tapaswi@kit.edu or fidler@cs.toronto.edu

Registration and submissions are open!
Benchmark: http://movieqa.cs.toronto.edu

Generic QA framework
- all multiple-choice QA approaches

Data representations
- represent data in vector space
  - TF-IDF facilitates matching exact words
  - Word2Vec meanings of words, allows synonyms
  - SkipThought encodes semantics of sentence

SS with Convolutional Brain
- core idea
  - search within the story to find the best match for question and the answer
  - windowed cosine similarity to compute how well a story fits a Q and A
  - answer questions, don’t look at story
  - pick correct answer as the
    - longest answer
    - most similar/distinct answer
    - answer most similar to the question

Evaluation
- plot-based answering easier, words repeated
- simple cosine similarity does not work with DVS, subtitles, scripts
- memory networks able to leverage this info.
- SSCB easily fuses all text representations

Method | Plot | DVS | Subtitle | Script
--- | --- | --- | --- | ---
Cosine TF-IDF | 47.6 | 24.5 | 24.5 | 24.6
Cosine Word2Vec | 46.4 | 26.6 | 24.5 | 23.4
Cosine SkipThought | 31.0 | 19.9 | 21.3 | 21.2
SSCB TF-IDF | 48.5 | 24.5 | 27.6 | 26.1
SSCB Word2Vec | 45.1 | 24.8 | 24.8 | 25.0
SSCB SkipThought | 28.3 | 24.5 | 20.8 | 21.0
SSCB Fusion | 56.7 | 24.8 | 27.7 | 28.7
MemN2N 1 layer | 40.6 | 33.0 | 38.0 | 42.3
MemN2N 3 layers | 42.3 | 33.0 | 37.1 | 43.0

The Hasty Student
- answer questions, don’t look at story
- pick correct answer as the
  - longest answer
  - most similar/distinct answer
  - answer most similar to the question

The Searching Student
- search within the story to find the best match for question and the answer
- windowed cosine similarity to compute how well a story fits a Q and A

The Hasty Turker
- 10 AMT workers answer questions without looking at the story

For all multiple-choice QA approaches

use a three-way function between story, question, and answers
e.g. a CNN-RNN approach on VQA
1. CNN(story = image); 2. RNN(question); 3. answer = softmax(vocabulary)

SS with Convolutional Brain
- core idea
  - search within the story to find the best match for question and the answer
  - windowed cosine similarity to compute how well a story fits a Q and A
  - answer questions, don’t look at story
  - pick correct answer as the
    - longest answer
    - most similar/distinct answer
    - answer most similar to the question

Evaluation
- plot-based answering easier, words repeated
- simple cosine similarity does not work with DVS, subtitles, scripts
- memory networks able to leverage this info.
- SSCB easily fuses all text representations

Method | Plot | DVS | Subtitle | Script
--- | --- | --- | --- | ---
Cosine TF-IDF | 47.6 | 24.5 | 24.5 | 24.6
Cosine Word2Vec | 46.4 | 26.6 | 24.5 | 23.4
Cosine SkipThought | 31.0 | 19.9 | 21.3 | 21.2
SSCB TF-IDF | 48.5 | 24.5 | 27.6 | 26.1
SSCB Word2Vec | 45.1 | 24.8 | 24.8 | 25.0
SSCB SkipThought | 28.3 | 24.5 | 20.8 | 21.0
SSCB Fusion | 56.7 | 24.8 | 27.7 | 28.7
MemN2N 1 layer | 40.6 | 33.0 | 38.0 | 42.3
MemN2N 3 layers | 42.3 | 33.0 | 37.1 | 43.0

The Hasty Student
- answer questions, don’t look at story
- pick correct answer as the
  - longest answer
  - most similar/distinct answer
  - answer most similar to the question

The Searching Student
- search within the story to find the best match for question and the answer
- windowed cosine similarity to compute how well a story fits a Q and A

The Hasty Turker
- 10 AMT workers answer questions without looking at the story

For all multiple-choice QA approaches

use a three-way function between story, question, and answers
e.g. a CNN-RNN approach on VQA
1. CNN(story = image); 2. RNN(question); 3. answer = softmax(vocabulary)