Multiple choice Q&A system

Kaggle competition

The Allen AI Science Challenge

Wed 7 Oct 2015 – Sat 13 Feb 2016 (7 months ago)

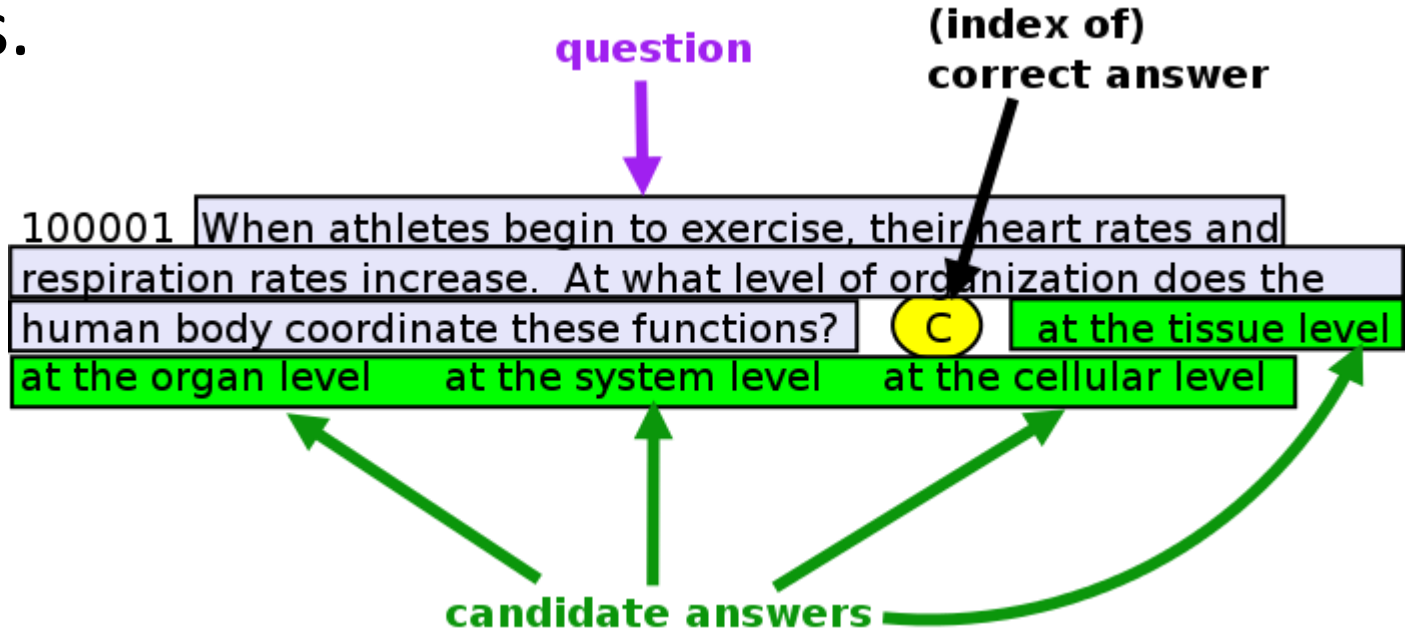
Competition Details » Get the Data » Make a submission

Is your model smarter than an 8th grader?

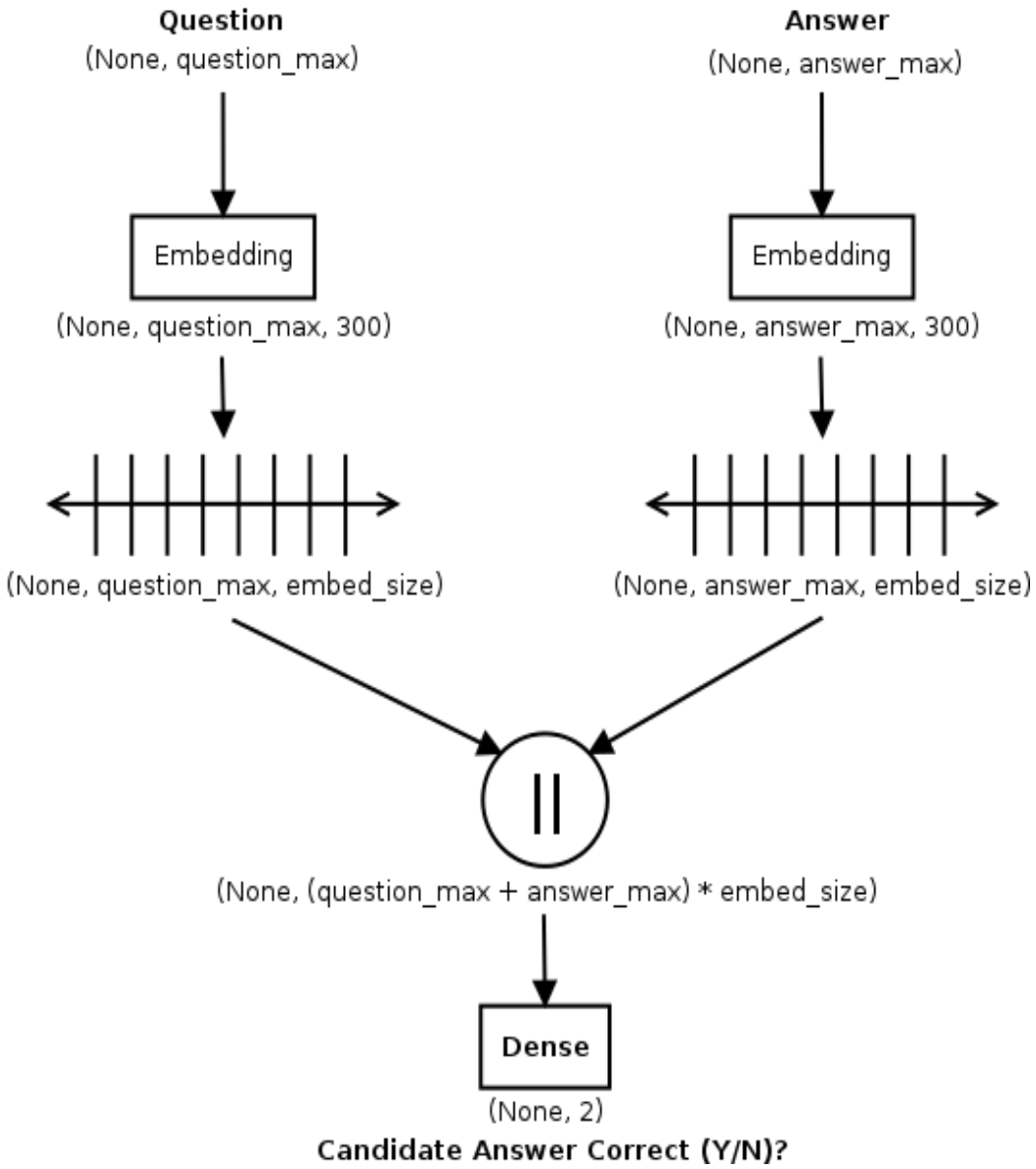
and the [#4 ranked entry] used Deep Learning for their solution.

Data Format

- Multiple choice 7^{th}/ 8^{th} Grade Science questions with 4 candidate answers and correct answer label.
- 2000 questions without correct answer label.
- Each question = 1 positive + 3 negative examples.

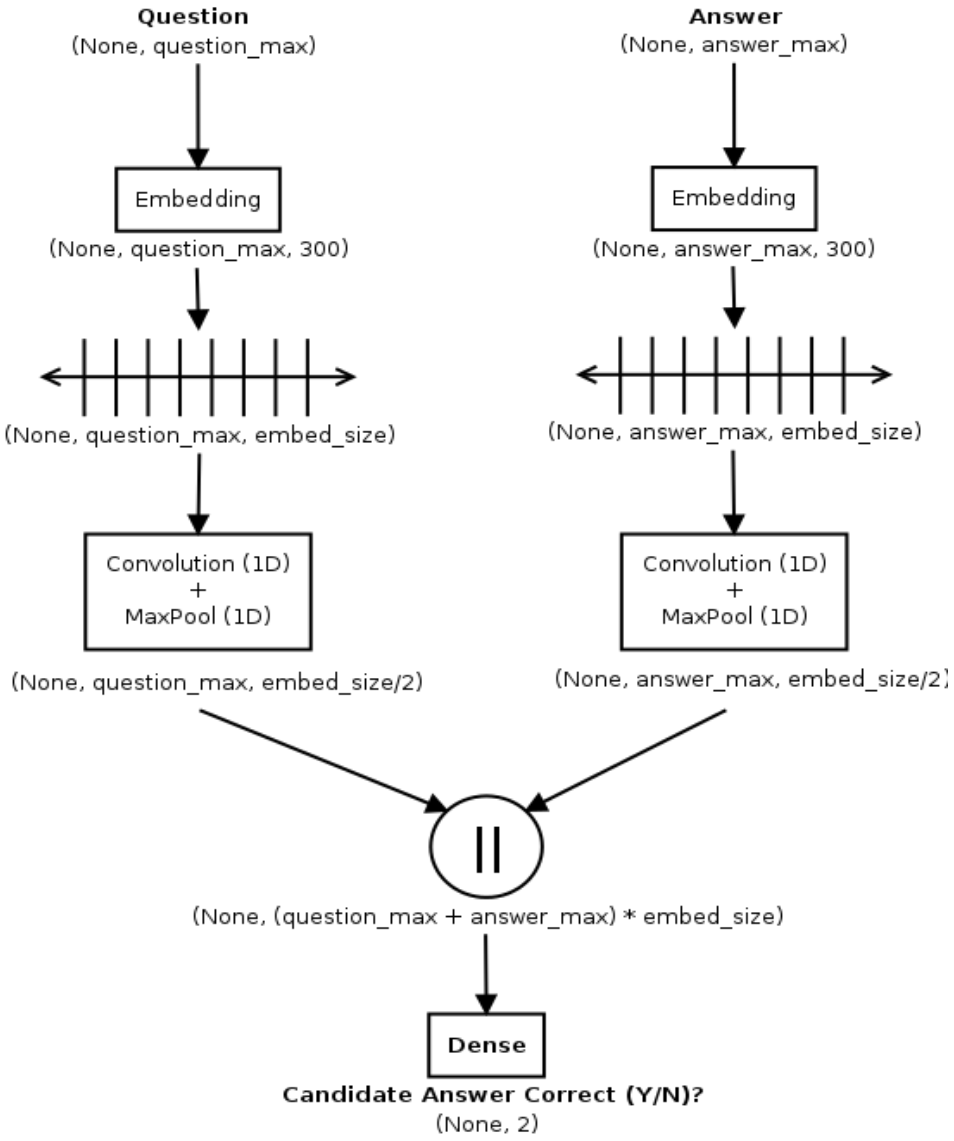


Approach1: LSTM



- Implementation based on the paper: **LSTM-based Deep Learning Models for Non-factoid Answer Selection**.
- Test accuracy reported in paper: 64.3%

Approach2: LSTM + CNN

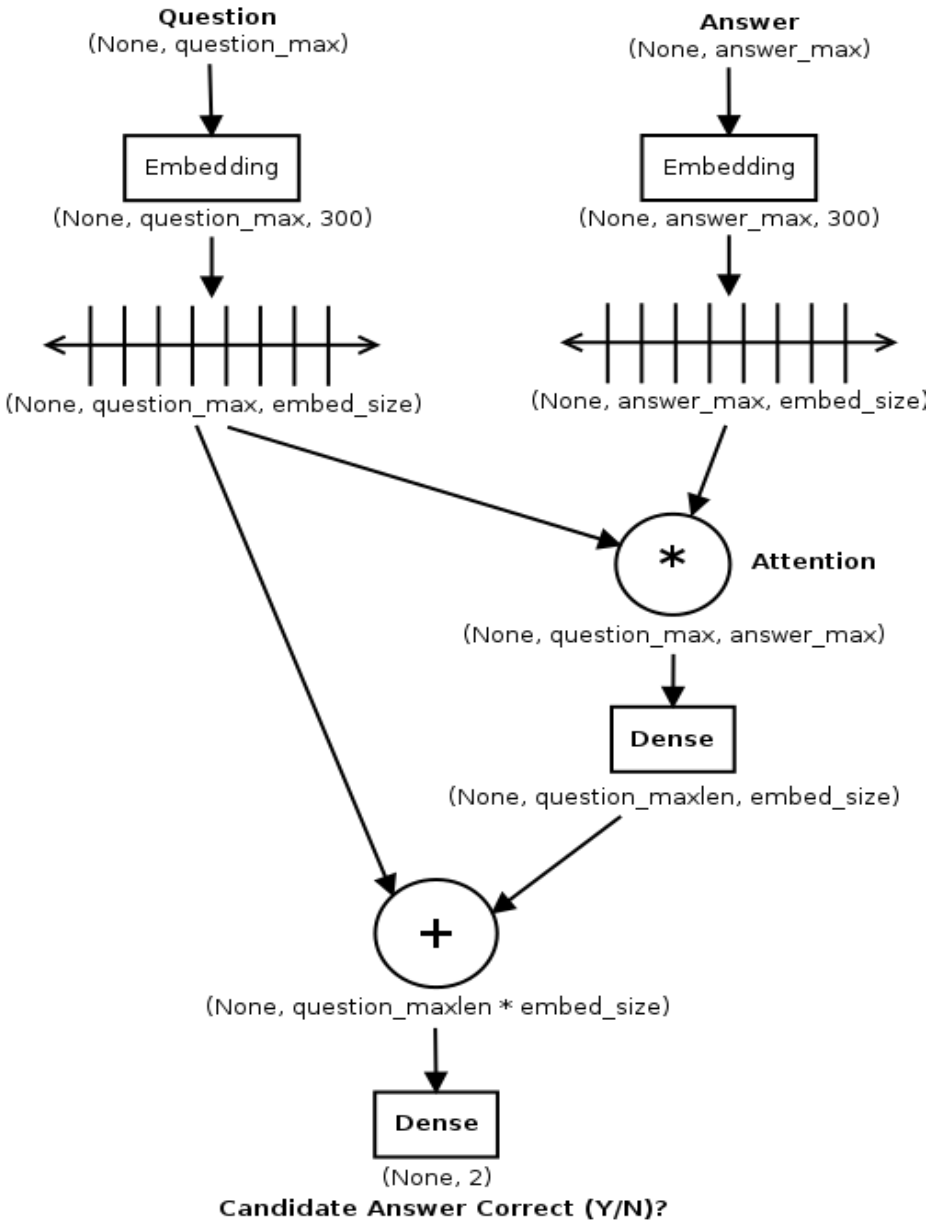


- Additional CNN Layer for more effective summarization.
- Test accuracy reported in paper: 62.2%

Incorporating Attention

- Vanishing Gradient problem addressed by LSTMs, but still shows up in long range Q+A contexts.
- Solved using Attention Models
  - Based on visual models of human attention.
  - Allow the network to focus on certain words in question with “high resolution” and the rest at “low resolution”.
  - Similar to advice given for comprehension tests about reading the questions, then scanning passage for question keywords.
  - Implemented here as a dot product of question and answer, or question and story vectors.

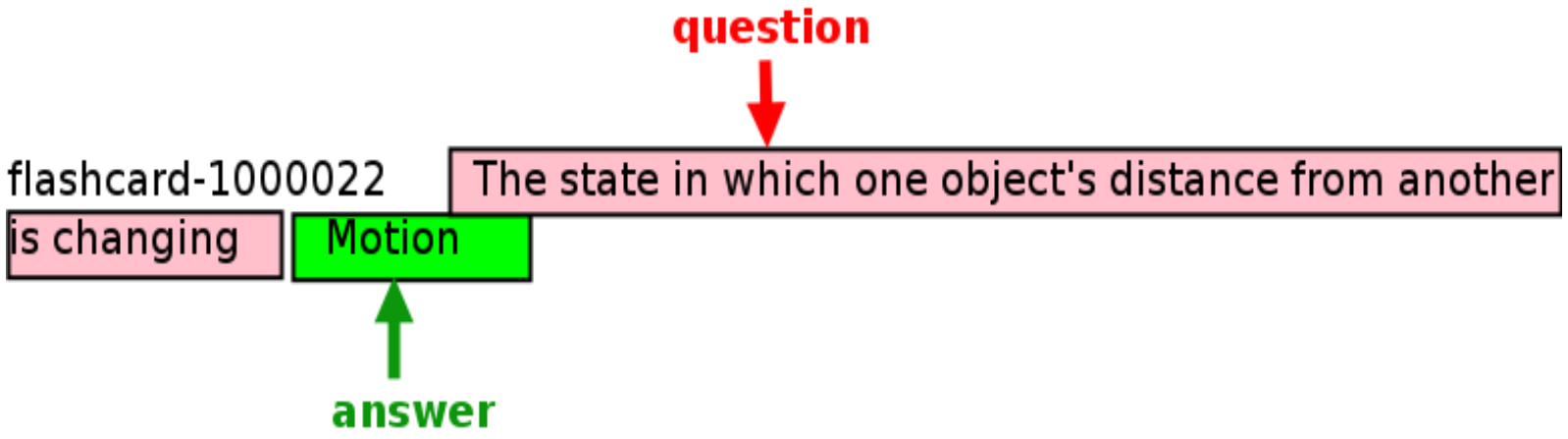
Approach3: LSTM + Attention



- Attention vector from question and answer combined with question.
- Test accuracy reported in paper: 68.4% (InsuranceQA dataset).

Incorporating External Knowledge

- Flashcard “story” = question || answer



Incorporating External Knowledge

- Contestants were allowed/advised to use external sources such as ConceptNet, CK-12 books, Quizlets, Flashcards from StudyStack, etc.
- Significant crawling/scraping and parsing effort involved. 4th place winner provides parsed download of StudyStack Flashcards on his Google drive.
- Flashcard "story" = question || answer

Using Story Embedding

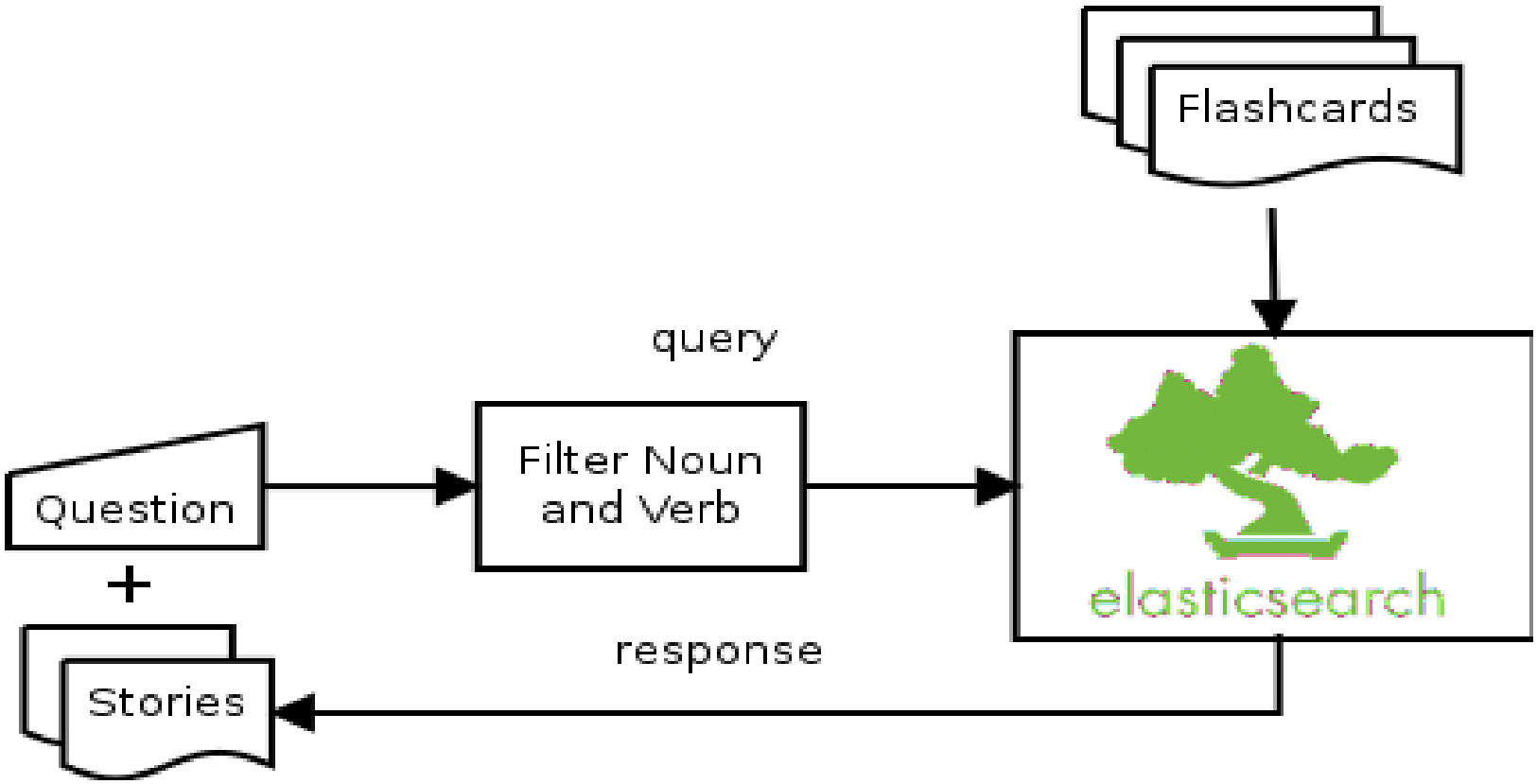
- Build Word2Vec model using words from Flashcards.
- Approximately 500k flashcards, 8,000 unique words.
- Provides smaller, more focused embedding space.
- Good performance boost over default Word2Vec embedding.

Model	Default Embedding	Story Embedding
QA-LSTM with Attention	62.93	76.27
QA-LSTM Bidirectional with Attention	60.43	76.27

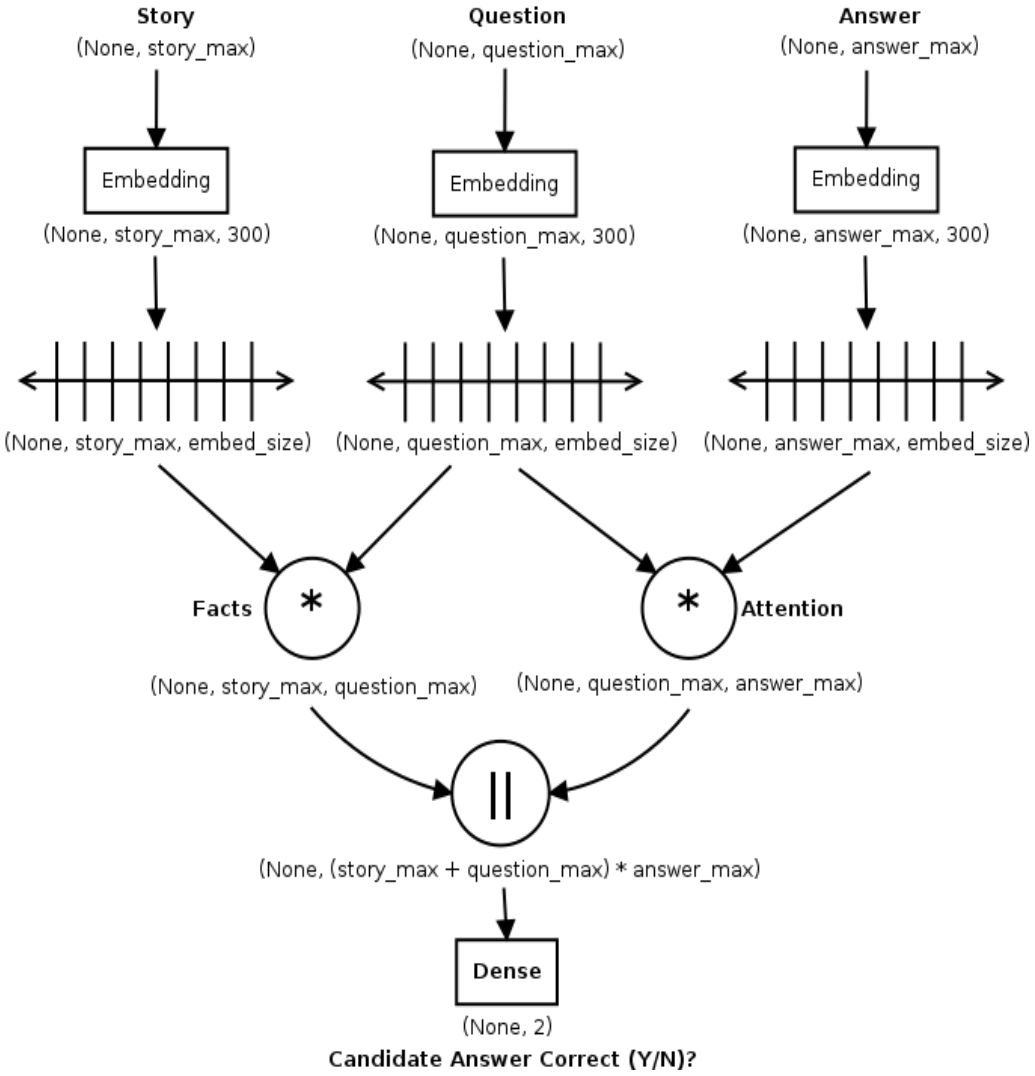
Relating Story to Question

- Replicate bAbI setup: (story, question, answer).
- Only a subset of flashcards relate to given question.
- Using traditional IR methods to generate flashcard stories for each question.

Relating Story to Question



Approach4: LSTM + Attention + Stories



- Story and Question combined to create Fact vector
- Question and Answer combined to create Attention vector
- Fact and Attention vectors concatenated

Model Deployment

- Our models predict answer correct vs. incorrect.
- Task is to choose the correct answer from candidate answers.
- Re-instantiate trained model with Softmax layer removed.
- Run batch of (story, question, answer) for each candidate answer.
- Select best scoring answer as correct answer.

Deploying Model - Example

Which is a distinction between an epidemic and a pandemic

[A] the symptoms of the disease
[B] the geographical area affected
[C] the species of organisms infected
[D] the season in which the disease spreads