Question Answering

Lecture Presentation

Ramya Rao Basava and Sang-Wha Sien

Outline

- Introduction to Question Answering
- Methods for Question Answering:
  - Information Retrieval based Methods
  - Knowledge based methods
  - Combined methods
  - Visual Question Answering
- Evaluation metrics
- Research Progress
  - Conversational Question Answering
- Summary

Note: Most of the slides in the presentation are based on Chapter 25 of Dan Jurafsky and James H. Martin. Speech and Language Processing, 3rd ed. draft. 2019, and https://github.com/oxford-cs-deepnlp-2017/lectures/blob/master/Lecture%2011%20-%20Question%20Answering.pdf , (including figures, text, references therein).

Introduction to QA

Question Answering (QA):

Discipline within the field of Natural Language Processing (NLP) that deals with building computer systems to answer human questions in a natural language.

IBM DeepQA system: Watson

QA Applications:

- QA has many popular applications and its methods are widely used in search, information extraction, summarization, dialogue in chats.
- A few examples of QA applications are:
  - IBM's Watson
  - Voice assistants like Google assistant or Apple's Siri
  - Pedagogical tools that answer student questions on reading comprehension tasks

The Google Assistant on the Pixel XL phone

Google assistant

Ref: https://en.wikipedia.org/wiki/Watson_(computer)

Ref: https://en.wikipedia.org/wiki/Google_Assistant

Information Retrieval based Methods

Information Retrieval (IR) based methods:

Best suited for:

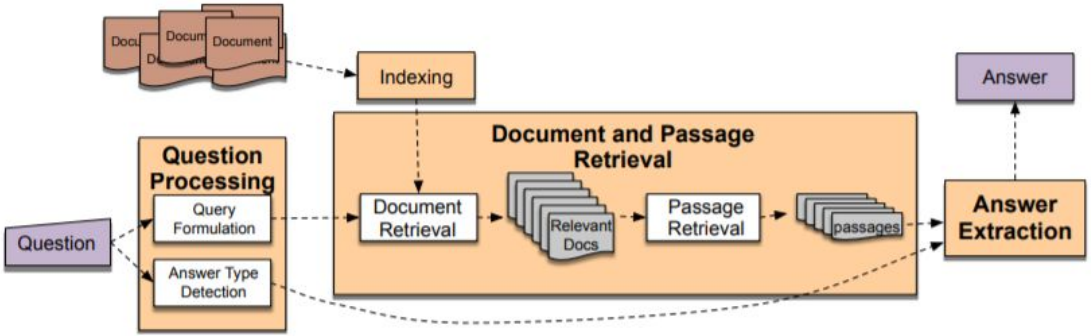
- Answering factoid questions
- Open domain information retrieval like the web

Question	Answer
Where is the Louvre Museum located?	in Paris, France
What's the abbreviation for limited partnership?	L.P.
What are the names of Odin's ravens?	Huginn and Muninn
What currency is used in China?	the yuan
What kind of nuts are used in marzipan?	almonds
What instrument does Max Roach play?	drums
What's the official language of Algeria?	Arabic
How many pounds are there in a stone?	14

Examples of factoid questions

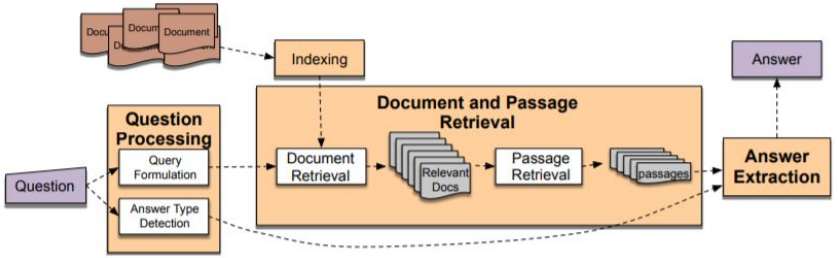
Generally carried out in three phases:

- Question Processing
- Document and Passage Retrieval
- Answer Extraction



Three phases of IR based systems

Information Retrieval based Methods



Question Processing:

Consists usually of the following tasks:

- Query formulation:
  - Need to create the query from the question, to a form appropriate for the document retrieval stage.
  - Usually minimal change is required for IR methods. For e.g., in web search we can directly use the input question.
  - Sometimes reformulation or rephrasing of the question is required, suitably applied to get the answer.
- Answer type detection:
  - Classifies the answer category
  - Usually based on methods like:
    - Building hierarchical answer type taxonomies from large lexical databases like WordNet.
    - Supervised learning where answer types are manually labeled for training.

Document and Passage Retrieval:

- A set of passages from documents or web are extracted by the QA system, which are passed to answer extraction phase.

Information Retrieval based Methods

Answer extraction:

Previous Methods:

- Baseline methods: like using a named entity tagger
- Feature-based supervised learning methods: the classifier is trained to recognize whether a sentence contains the answer.

Modern Methods:

Using Neural Networks (NN): Basic assumption is that the question and the answer are semantically similar in some way. Generally have the following tasks:

- Embedding for the question and words of the passage are created
- Spans of the passage which is most similar to the question are selected as answer spans.

Popular NN methods for QA:

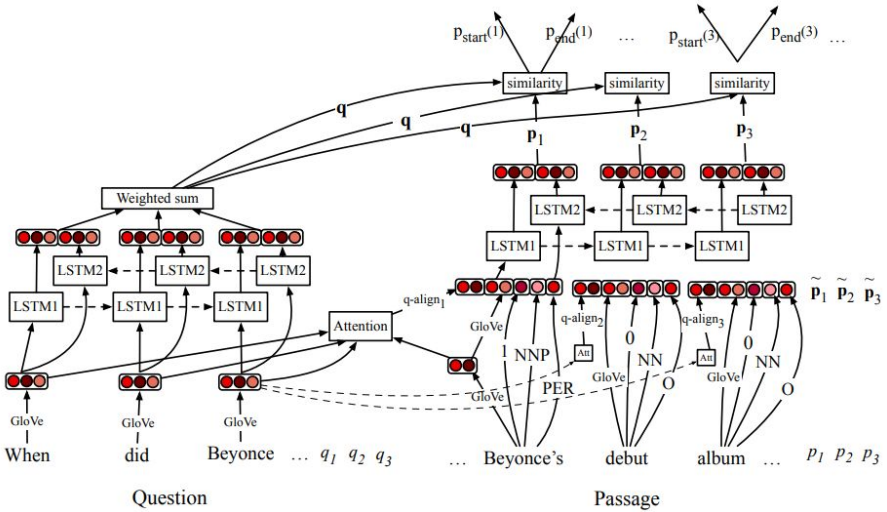
- A bi-LSTM-based Reading Comprehension Algorithm
- BERT-based Question Answering

Information Retrieval based Methods

Answer extraction

A bi-LSTM-based Reading Comprehension Algorithm:

- A combined question embedding is created by passing the initial GLoVe embeddings of the words in the question through a bi-LSTM network.
- Similarly embedding for the passage words is created by:
  - Concatenating a set of pre-processing embedding steps to create the initial embedding, which is then passed through a bi-LSTM network.
- Start and end probabilities of each word in the passage are obtained by comparing its similarity with the question, in the downstream task.
- Words with the highest start and end probability constitute the answer span.

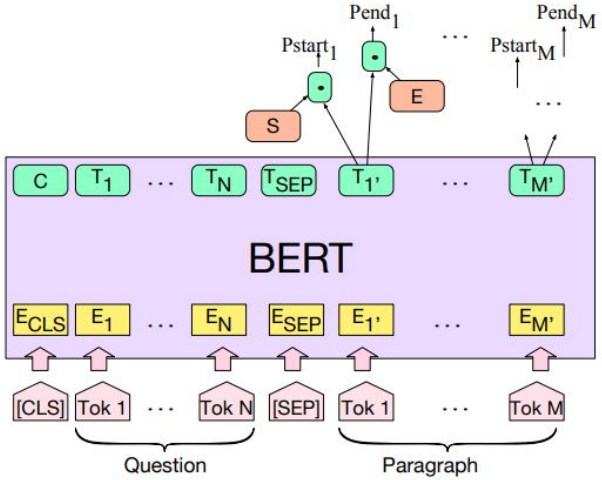


Information Retrieval based Methods

Answer extraction

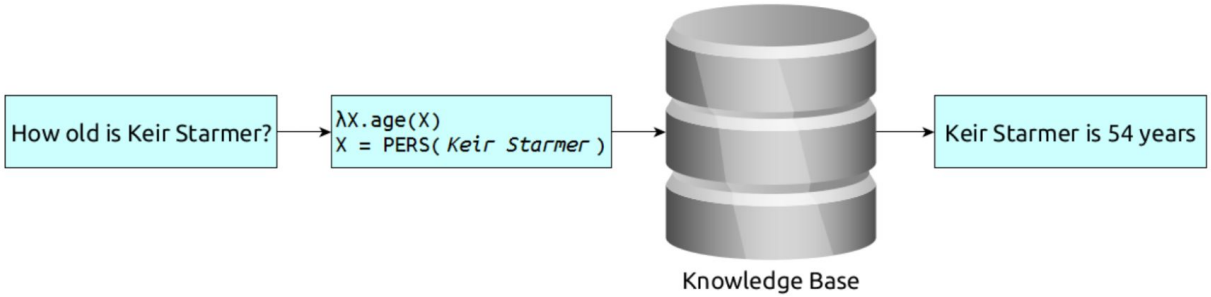
BERT-based Question Answering:

- For the QA task, the two input strings constitute the question and the passage, separated by SEP token as shown in the image.
- From the output tokens for the passage for the BERT model, we can obtain the probabilities for the words to be a start-span word or end-span word for the answer, by training with some similarity measure with respect to the question in the fine-tuning phase.
- BERT is one of the landmark methods in NLP and provides excellent results for tasks like QA.



Knowledge Based (KB) Methods

- Why Knowledge Based?
  - Much the available information is stored in structured formats (e.g. relational databases)
- We assume that Logical Form → KB Query → Answer is easy to do
- All that is left to do is the Question→Logical Form step.



Question → Logical Form → KB Query → Answer

Knowledge Based (KB) Methods

Semantic Parsing is the process of mapping natural language into a formal representation of its meaning. Depending on the chosen formalism (e.g. lambda expressions, query language), this logical representation can be used to query a structured knowledge base.

Question	Logical form
When was Ada Lovelace born?	`birth-year (Ada Lovelace, ?x)`
What states border Texas?	λ x.state(x) $\wedge$ borders(x,texas)
What is the largest state	argmax(λx.state(x), λx.size(x))
How many people survived the sinking of the Titanic	`(count (!fb:event.disaster.survivors fb:en.sinking_of_the_titanic))`

"Where are kindergartens in Hamburg?"

```
query(area(keyval(name,Hamburg)),
      nwr(keyval(amenity,kindergarten)),
      qtype(latlong))
```

Knowledge Based (KB) Methods

Three different types of KB methods

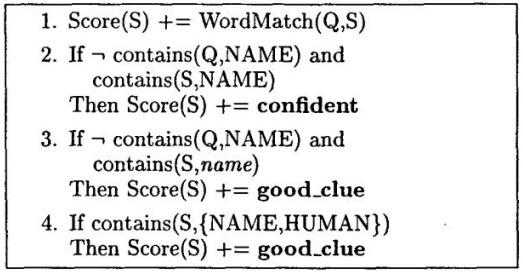
1. **Rule-based methods** use heuristic rules, often handwritten, in order to find lexical or semantic information in the question and document.
2. **Supervised methods** uses training data that contain questions paired with their correct logical forms.
3. **Semi-supervised and Unsupervised methods** need to construct the logical forms, often using redundant information in the web.

Knowledge Based (KB) Methods

Rule-based methods example: Quarc system

Riloff and Thelen (2000): A rule-based question answering system for reading comprehension tests.

- Uses different rules based on WH question type (who, what, when, where, why) to identify a span of text that holds the answer in a reading comprehension task.
- Rules can either be lexical (matching words in the question and in the sentence) or semantic (recognizing semantic classes, e.g. HUMAN, NAME).
- After identifying the question type, rules are applied to each sentence of the context document and points are awarded on how strongly a rule applies to a sentence.
- The sentence with the highest score is returned as the answer to the given question



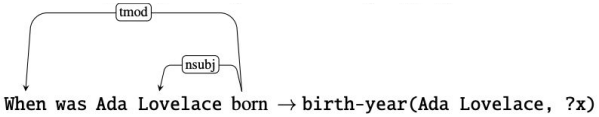
WHO rules

Knowledge Based (KB) Methods

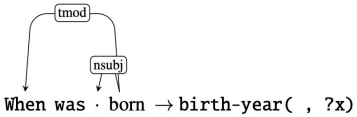
Supervised methods example

Zettlemoyer and Collins (2005): Learning to map sentences to logical form: structured classification with probabilistic categorial grammars.

- Can learn from given question-logical form pairs to construct generalized rules that can map unseen questions to logical forms.
- The learning process involves parsing the questions and aligning the parse trees to the logical form.



- The mapping of similar but unseen questions to their respective logical forms can be made by generalizing the above parsing.



- Extensions
  - Can use probabilities to choose the best parsing for each sentence
  - For more complex questions, break them down into smaller more manageable units and combine them later.

Knowledge Based (KB) Methods

Unsupervised methods example: REVERB

Fader et al. (2011): Identifying relations for open information extraction

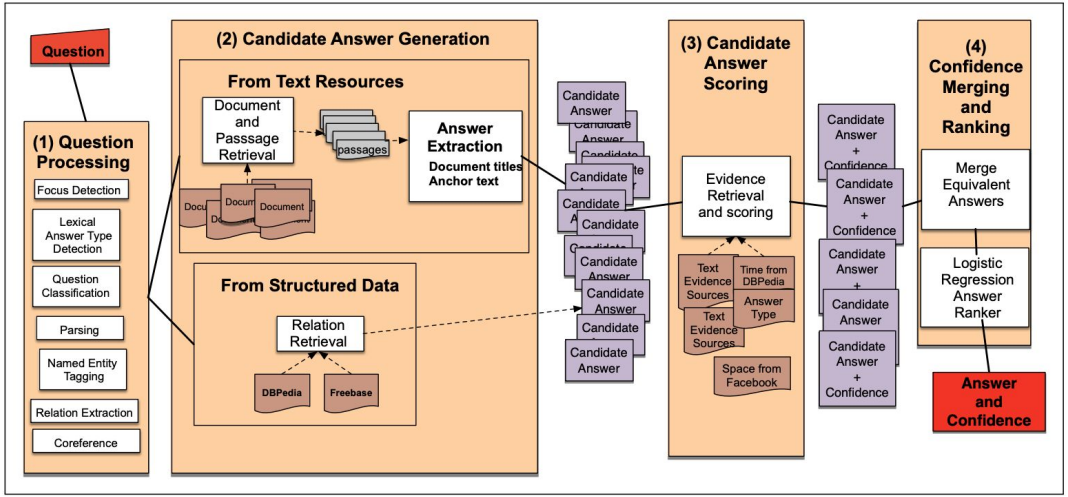
- The problem with supervised methods is that simple relational forms cannot cover the wide variety of questions that can be asked. And it is very expensive to annotate because human experts are needed to create the logical forms.
- Solution: can take advantage of redundancy in the text to structure the logical forms, making use of the vast amount of information found in the web. This can be done in a semi-supervised or unsupervised way.
- Creates subject-relation-object triple strings (e.g. “Ada Lovelace”, “was born in”, “1815”)
- Go through Wikipedia to extract similar strings (cosine distance) using a predicate.
- The resulting set of strings can be consulted to map more variety of questions to relations.
- Results for the Freebase relation of *country.capital*:

capital of	capital city of	become capital of
capitol of	national capital of	official capital of
political capital of	administrative capital of	beautiful capital of
capitol city of	remain capital of	make capital of
political center of	bustling capital of	capital city in
cosmopolitan capital of	move its capital to	modern capital of
federal capital of	beautiful capital city of	administrative capital city of

Combined Methods: IBM Watson

Ferrucci et al. (2010): Building Watson: An Overview of the DeepQA Project

- Uses a combined approach
  - Generates multiple candidate answers from both text resources and structured data (2)
- Question processing much like IR methods (1)
- Candidate answers are scored (3) and similar answers are merged (e.g. JFK, John F. Kennedy) (4)
- Learns the probabilities of the answers being correct and chooses the best answer with the highest probability (4)



IBM Watson

Answer Types in *Jeopardy!*

- 2500 answer types in 20,000 Jeopardy question sample
- The most frequent 200 cover < 50% of the data
- The 40 most frequent Jeopardy answer types:

he, country, city, man, film, state, she, author, group, here, company, president, capital, star, novel, character, woman, river, island, king, song, part, series, sport, singer, actor, play, team, show, actress, animal, presidential, composer, musical, nation, book, title, leader, game

Visual QA

Visual Question Answering:

- VQA is a multi-disciplinary area of research: combines Computer Vision (CV) and NLP, where systems answer questions in natural language.
- Task: Answer a query question on the image, similar to how humans interpret the image.
- Goal is to achieve open-ended free-form question answering from images.

Applications:

- AI systems can extract/provide more information to queries, from images.
- For e.g. useful for visually challenged people.

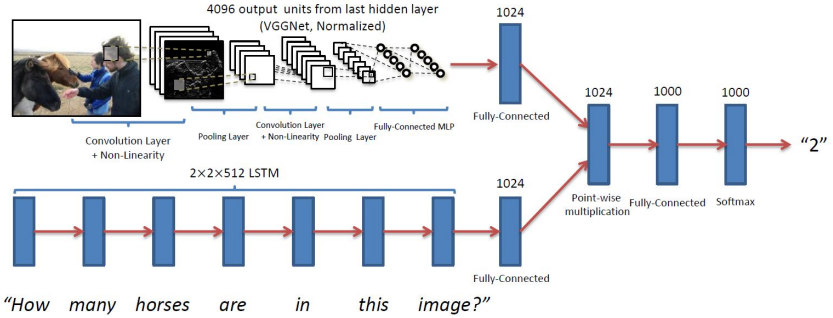
Popular method:

- VQA: Visual Question Answering (Agrawal et al. 2015)

Visual QA

VQA: Visual Question Answering (Agrawal et al. 2015)

- This paper proposes free-form and open-ended VQA, where given an image and a question related to that image, the computer system should be able to provide an answer in natural language.
- This method also takes into account knowledge based reasoning.
- A deep neural network model which combines both image processing and NLP is proposed.
- A deep convolutional neural network (VGGNet) is used for image processing and a LSTM based network is used for natural language processing of the question.
- Output from the two channels above, gives a combined image + question embedding which is passed through a softmax to give a set of answers.
- Answer with the highest activation is chosen as the final answer.



Ref: VQA: Visual Question Answering (Agrawal et al. 2016)

Ref: Very Deep Convolutional Networks for Large-Scale Image Recognition (Simonyan et al. 2014)

Demos

Google AI on Natural Questions corpus
https://ai.google.com/research/NaturalQuestions/visualization

VQA
https://vqa.cloudcv.org/

BERT QA
https://www.pragnakalp.com/demos/BERT-NLP-QnA-Demo/

Microsoft QA system
https://www.microsoft.com/en-us/research/welcome-to-canada-demo/

BIDAF on SQuAD dataset
http://allgood.cs.washington.edu:1995/

QA Evaluation Metrics

- Mean Reciprocal Rank (MRR)
  - Assumes systems return a ranked list of answers

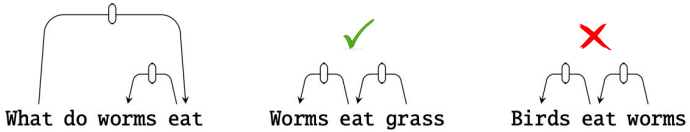
$$\text{MRR} = \frac{1}{N} \sum_{i=1 \text{ s.t. } rank_i \neq 0}^{N} \frac{1}{rank_i}$$

- F1 Score - Average overlap between predicted and gold answers
- Exact Match (Accuracy) - % predicted answers that match the gold answer
- Perplexity - for neural approaches
- BLEU Score - Machine Translation measure for translation accuracy
  - Bilingual Evaluation Understudy (Papineni 2002)
  - Machine translation methods are often used in unsupervised methods to transform questions to logical forms.

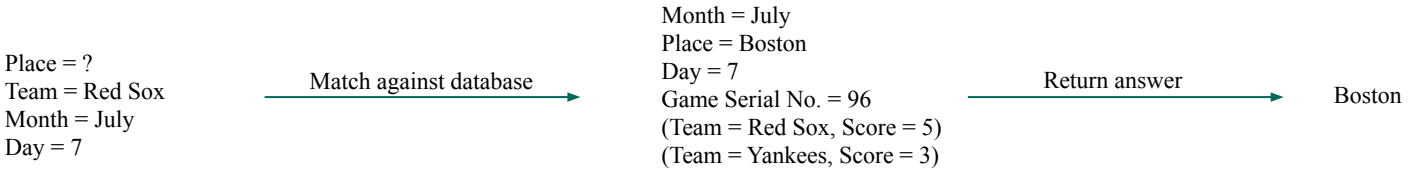
Research Progress

History

- Text-based paradigm: Protosynthex
  - Simmons et al. (1964): Indexing and dependency logic for answering english questions
  - First forms a query from the content words in the question, and then retrieves candidate answer sentences in the document, ranked by their frequency-weighted term overlap with the question.
  - The query and each retrieved sentence are parsed with dependency parsers, and the sentence whose structure best matches the question structure is selected.



- Knowledge-based paradigm: BASEBALL system
  - Green et al. (1961): Baseball: An automatic question answerer
  - Queries a structured database of game info (i.e. value-attribute pairs of each game)
  - Question: Where did the Red Sox play on July 7?



Research Progress

Current Research

- http://nlpprogress.com/english/question_answering.html
  - List of QA research and datasets as well as current state of the art
  - At least 25 different datasets across 3 broad types of applications: reading comprehension, knowledge base, open domain
- Reading comprehension
  - Conversational QA (QUAC, CoQA)
  - Cloze-style reading comprehension (CliCR, CNN/Daily Mail, Story Cloze Test)
  - Commonsense inference sentence completion (SWAG, CODAH)
  - Inference from multiple sentences or documents (DuoRC, WikiHop, MedHop)
- Reading comprehension state of the art
  - Leaderboards show that state of the art are mostly neural, specifically systems that feature attention mechanisms
    - BiDAF - Seo et al. (2016): Bidirectional Attention Flow for Machine Comprehension
    - BERT models are among the state of the art for many datasets: CoQA, QUAC, SWAG, SQuAD 2.0
      - Lan et al. (2019): ALBERT: A Lite BERT for Self-supervised Learning of Language Representations
      - Liu (2019): RoBERTa: A Robustly Optimized BERT Pretraining Approach

Research Progress

Current Research (continued)

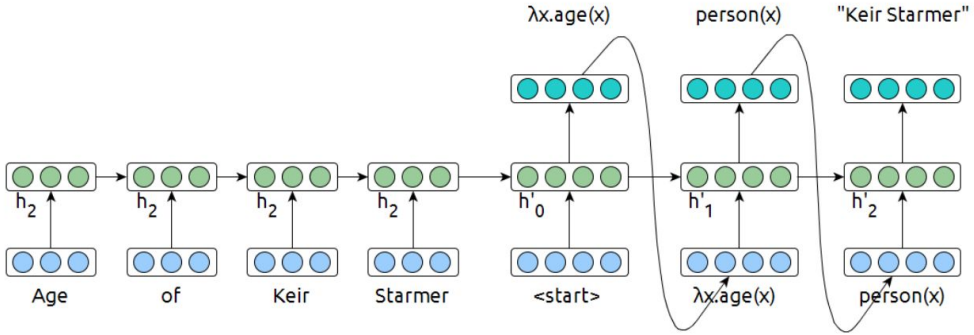
- Knowledge Based QA
  - Neural Semantic Parsing
    - Like in machine translation, using attention can be helpful
      Dong and Lapata (2016): Language to Logical Form with Neural Attention
    - Exploit the highly rigid structure in the target side to constrain generation
      Liang et al. (2016): Neural Symbolic Machines
      Ling et al. (2016): Latent predictor networks for code generation
    - Make use of semi-supervised training to counter sparsity
      Kocisky et al. (2016): Semantic Parsing with Semi-Supervised Sequential Autoencoders

Research Progress

Current Research (continued)

A deep learning approach to semantic parsing

- Semantic parsing can be viewed as a sequence to sequence model, similar to machine translation.
- Encode sentence with sequence models
- Decode with standard mechanisms from MT
- But…
  - Supervised training data hard to come by
  - Depending on formalism used, highly complex target side
  - How to deal with proper nouns and numbers?

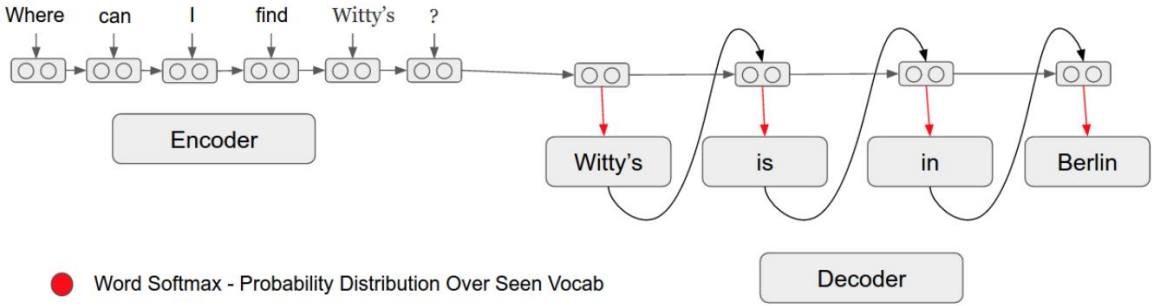


Research Progress

Current Research (continued)

Generation with multiple sources

- Ling et al. (2016): Latent predictor networks for code generation

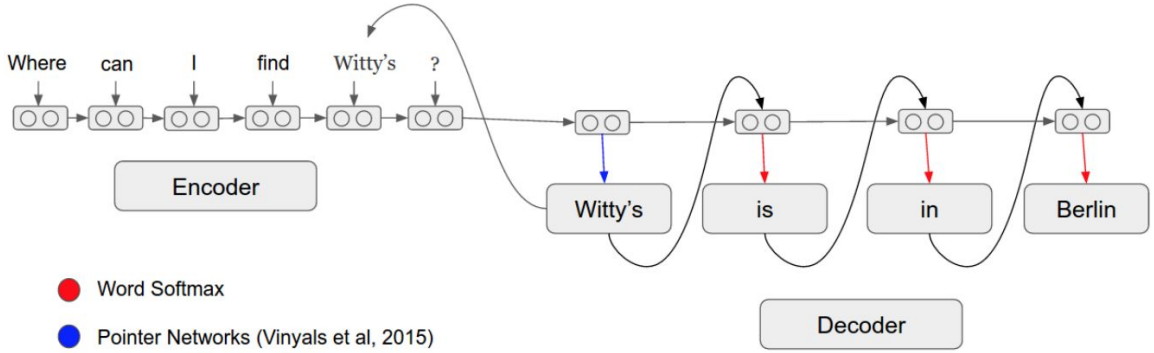


Research Progress

Current Research (continued)

Generation with multiple sources

- Ling et al. (2016): Latent predictor networks for code generation

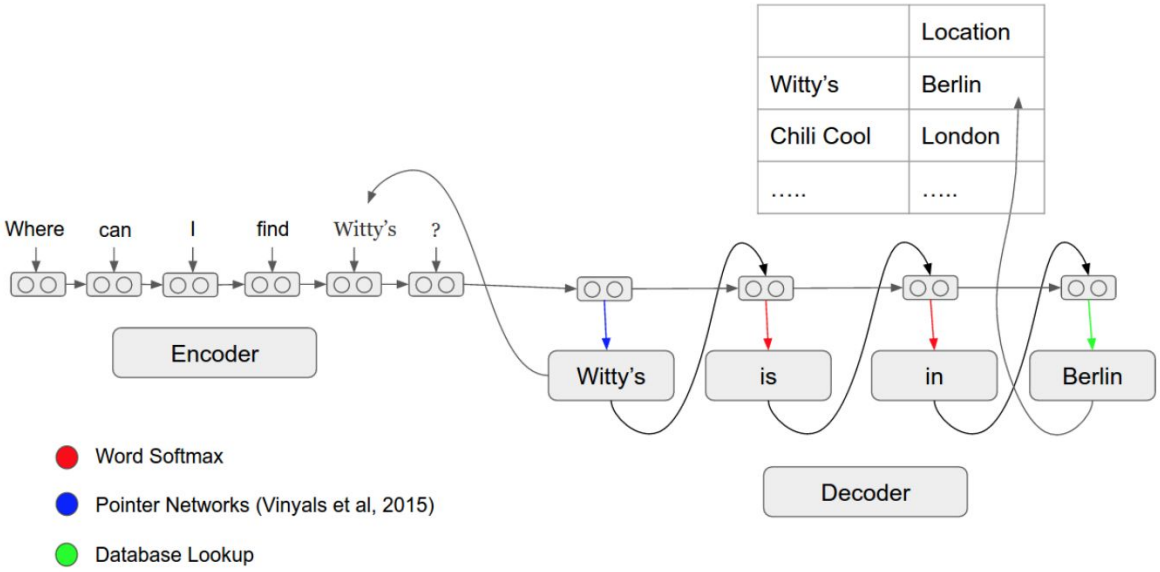


Research Progress

Current Research (continued)

Generation with multiple sources

- Ling et al. (2016): Latent predictor networks for code generation



Conversational QA (CoQA)

- https://stanfordnlp.github.io/coqa/
- Large-scale dataset from Stanford NLP group
- Task: measure the ability of machines to understand a span of text and participate in a dialogue
- Motivation: people don't ask just one question to learn a topic but a series of questions to follow up.
- But using conversation history is difficult for computers because co-references are ambiguous.
- Co-references: when multiple expressions refer to the same thing
  - Explicit co-references: pronouns
  - Implicit co-references: one word questions
- CoQA is filled with co-references
- State of the art:
  - Ju et al. (2019): Technical report on Conversational Question Answering
  - Uses RoBERTa + AT + KD based on BERT, Adversarial Training, and Knowledge Distillation

The Virginia governor's race, billed as the marquee battle of an otherwise anticlimactic 2013 election cycle, is shaping up to be a foregone conclusion. Democrat Terry McAuliffe, the longtime political fixer and moneyman, hasn't trailed in a poll since May. Barring a political miracle, Republican Ken Cuccinelli will be delivering a concession speech on Tuesday evening in Richmond. In recent ...

Q_1: What are the candidates **running** for?
A_1: Governor
R_1: The Virginia governor's race

Q_2: **Where**?
A_2: Virginia
R_2: The Virginia governor's race

Q_3: Who is the democratic candidate?
A_3: **Terry McAuliffe**
R_3: Democrat Terry McAuliffe

Q_4: Who is **his** opponent?
A_4: **Ken Cuccinelli**
R_4 Republican Ken Cuccinelli

Q_5: What party does **he** belong to?
A_5: Republican
R_5: Republican Ken Cuccinelli

Q_6: Which of **them** is winning?
A_6: Terry McAuliffe
R_6: Democrat Terry McAuliffe, the longtime political fixer and moneyman, hasn't trailed in a poll since May

https://towardsdatascience.com/coqa-a-conversational-question-answering-challenge-7d9c3aceda3c

Summary

- QA is a discipline within the field of Natural Language Processing (NLP) that deals with building computer systems to answer human questions in a natural language
- Two major paradigms in QA:
  - Information retrieval (IR) based methods: used to answer factoid questions
    - Question Processing, Document and Passage Retrieval, Answer Extraction
  - Knowledge based methods: rely on information encoded in structured formats
    - Rule-based, Supervised, Semi-supervised, Unsupervised methods
  - Combined approaches - IBM Watson
- QA is a rich area of research
  - Neural approaches tend to have better results (e.g. BERT models)
  - Datasets for wide range of applications - VQA, CoQA