Tutorial on Deep Learning for Natural Language Processing
ICON-2017, Jadavpur University, Kolkata, India.

Question Answering

Md Shad Akhtar
Research Scholar
AI-NLP-ML Group
Department of Computer Science & Engineering
Indian Institute of Technology Patna

shad.pcs15@iitp.ac.in
https://iitp.ac.in/~shad.pcs15/

Thanks to Deepak Gupta,
Research Scholar, IIT Patna
Question Answering

- System that automatically answer questions posed by humans in natural language.
Why Question Answering?

● Conversational Agents  
  ○ Ask it questions. Tell it to do things. Facebook (M), Apple (Siri), Google Assistant

● Biomedical and clinical QA  
  ○ There is a need of system that accepts the queries from medical practitioners in natural language and returns the answers quickly and efficiently.

● E-commerce sites  
  ○ E-commerce sites can exploit the user reviews to provide the answers of various user questions.
Categories of QA

● Factoid Questions  
  ○ Question that can be answered with simple facts expressed in short text answers.  
    ■ **Question:** Who is the author of the book Wings of Fire?  
    ■ **Answer:** A. P. J, Abdul Kalam

● List Questions  
  ○ Question requires multiple facts to be returned in answer to a question.  
    ■ **Question:** What are the islands in India?  
    ■ **Answer:** Andaman Island, Nicobar Island, Labyrinth Island, Barren Island

● Descriptive Questions  
  ○ The answer can be short descriptive from a single sentence to multiple but limited (2-3) sentences. It can also be the long descriptive where answer can be a paragraph with meaning full information of the natural language query.  
    ■ **Question:** What is Greenhouse effect?  
    ■ **Answer:** The analogy used to describe the ability of gases in the atmosphere to absorb heat from the earth’s surface.
A typical IR based QA

- **Question Processing**
  - Type of entity the answer should consist of (person, location, time, etc.).
  - The query specifies the keywords that should be used for the IR system to use in searching.

- **Passage Retrieval**
  - Collect the document against the query for documents.

- **Answer Processing**
  - To extract a specific answer from the passage.
  - Use information about the expected answer type to find the answer.
Sub-problems of QA

- **Semantic Question matching**
  - Given a natural language question $Q$ and set of candidate questions $CQ$, the task is to rank each of the question $Q_{cq} \in CQ$ according to their semantic similarity to the question $Q$,
    - Q1: *What are the best ways to lose weight?*
    - Q2: *How can a person reduce weight?*
    - Q3: *What are the effective weight loss plans?*

- **Answer Triggering**
  - Given a question and a set of answer candidates, **answer triggering** determines whether the candidate set contains any correct answers. If yes, it then outputs a correct one.
    - Q: *How big is BMC software in houston?*
    - A1: *BMC Software Inc. is an American company specializing in Business Service Management (BSM) software.*
    - A2: *Employing over 6,000, BMC is often credited with pioneering the BSM concept as a way to help better align IT operations with business needs.*
Problem formulation: Semantic Question Matching / Answer Triggering

● Let
  ○ Question is \( q \)
  ○ Candidate questions are \( CQ = \{cq_1, cq_2, \ldots, cq_n\} \)
  ○ Candidate answers are \( CA = \{ca_1, ca_2, \ldots, ca_n\} \)

● Question - Question pairing
  ○ \( <q,cq_1>, <q,cq_2>, \ldots, <q,cq_n> \)

● Question - Answer pairing
  ○ \( <q,ca_1>, <q,ca_2>, \ldots, <q,ca_n> \)

● Decision
  ○ Binary - Yes/No
  ○ Similarity score - a continuous value in the range 0 to 1.
Detecting Duplicate Questions with Deep Learning - Siamese Network

Yushi Homma, Stuart Sy and Christopher Yeh; Detecting Duplicate Questions with Deep Learning; In Proceedings of the 30th Conference on Neural Information Processing Systems (NIPS 2016), Barcelona, Spain;
Learning Hybrid Representations to Retrieve Semantically Equivalent Questions

Cícero dos Santos, Luciano Barbosa, Dasha Bogdanova, Bianca Zadrozny; Learning Hybrid Representations to Retrieve Semantically Equivalent Questions; In proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing, pages 694–699, Beijing, China, 2015.
Learning Hybrid Representations to Retrieve Semantically Equivalent Questions

\[
s_{bow}(q_1, q_2) = \frac{r_{bow}^{q_1} \cdot r_{bow}^{q_2}}{\|r_{bow}^{q_1}\| \cdot \|r_{bow}^{q_2}\|}
\]

\[
s_{conv}(q_1, q_2) = \frac{r_{conv}^{q_1} \cdot r_{conv}^{q_2}}{\|r_{conv}^{q_1}\| \cdot \|r_{conv}^{q_2}\|}
\]

\[
s(q_1, q_2) = \beta_1 \cdot s_{bow}(q_1, q_2) + \beta_2 \cdot s_{conv}(q_1, q_2)
\]
SelQA: A New Benchmark for Selection-based Question Answering
SelQA: A New Benchmark for Selection-based Question Answering

- For all $w \in T$, where $T$ is common words in Q & A.
  - $P_q$ & $P_a \rightarrow$ parents of $w$ in $D_q$ & $D_a$
  - $S_q$ & $S_a \rightarrow$ siblings of $w$ in $D_q$ & $D_a$
  - $C_q$ & $C_a \rightarrow$ children of $w$ in $D_q$ & $D_a$

- Three features: $f(P_q, P_a), f(S_q, S_a)$ and $f(C_q, C_a)$

Tomasz Jurczyk, Michael Zhai, Jinho D. Choi; SelQA: A New Benchmark for Selection-based Question Answering; In proceedings of the 2016 IEEE 28th International Conference on Tools with Artificial Intelligence (ICTAI), pages 820–827, San Jose, CA, USA, 2016...
Thank You!

AI-NLP-ML Group, Department of CSE, IIT Patna (http://www.iitp.ac.in/~ai-nlp-ml/)

Research Supervisors:

- Prof. Pushpak Bhattacharyya
- Dr. Asif Ekbal
- Dr. Sriparna Saha