MSE | MASTER OF SCIENCE IN ENGINEERING

Hes·so
Haute Ecole Spécialisée de Suisse occidentale
Fachhochschule Westschweiz
University of Applied Sciences and Arts Western Switzerland

Master of Science HES-SO in Engineering
Av. de Provence 6
CH-1007 Lausanne

Master of Science HES-SO in Engineering

Orientation: Information and Communication Technologies (ICT)

GraphQA, a Deep Retrieval Chatbot

A Multi-hop Conversational Question-Answering Chatbot using Sub-Knowledge Graphs

Author:

Romain Claret

Under the direction of:
Prof. Dr. Jean Hennebert
HES-SO//Fribourg
Institute of Complex Systems (iCoSys)

External expert:
Prof. Dr. Michael Ignaz Schumacher
HES-SO//Valais
Institute of Business Information Systems

Fribourg, HES-SO//Master, January 30, 2020

Abstract

We propose an innovative approach for question-answering chatbots to handle conversational contexts and generate natural language sentences as answers. In addition to the ability to answer open-domain questions, our zero-shot learning approach, which uses a pure algorithmic orchestration, provides a modular architecture to swap statically or dynamically task-oriented models while preserving its independence to training.

In the scope of this research, we realize the Proof-of-Concept of an Open-domain and Closed-ended Question-Answering chatbot able to output comprehensive Natural Language generated sentences using the Wikidata Knowledge Base.

To achieve the concept, we explore the extraction, and the use of sub-knowledge graphs from the Wikidata knowledge base to answer questions conversationally and to use the sub-graphs as context holder. Additionally, we are extracting Subject-Predicate-Object tuples from the graph and using Language Models to join the SPOs and extend the answers as natural language sentences.

The proof-of-concept architecture uses a combination of state-of-the-art and industry-used models with a fine-tuning strategy. As a motivational target, we use a Zero-Shot Learning approach, by combining various models with an algorithmic orchestrator and using pure algorithmic for the graph manipulation and answer extraction.

Finally, we evaluate the answers and compare the results with state-of-the-art Single-Hop and Multi-Hop question-answering systems on question-answering datasets. We find out that, aside from the computation time and the computational resources needed, our proof-of-concept performs similarly at question-answering compared to its competitors.

Keywords: Machine Learning (ML), Natural Language Processing (NLP), Single-Hop, Multi-Hop, Question Answering (QA), Wikidata, Wikipedia, Knowledge Graph (KG), Knowledge Base (KB), Word Embedding, Part of Speech Tagging, Named-Entity Recognition, Named-Entity Linking, Language Model (LM), Model Fine-Tuning, Graphs, Sub-Knowledge Graphs, Transformer, Bidirectional Encoder Representations from Transformers (BERT), Generative Pre-Training 2 (GPT-2), Information Extraction (IE), Spacy, GloVe, DeepCorrect, Chatbot, Conversational, Information Retrieval (IR), Queries, Python

Glossary

Adversarial Learning

In Machine Learning (ML), the concept of this technique relies on trying to fool models via malicious inputs. It can be interpreted as a game a model is playing with itself by modifying the input in such a way that the model will recognize it as another input then learn from its mistake.

Attention Mechanism

In Natural Language Processing (NLP), the Attention Mechanism is an algorithm used to calculate the relational weight between elements in a sequence of elements (most often words).

Close-ended

A closed-ended question is designed to allow a limited amount of responses.

Encoder-Decoder

In Machine Learning (ML), Encoder-Decoder is two Neural Networks (NNs) that work in pair. The Encoder generates a fixed-size output vector from any sized vector input. And the Decoder generates from the Encoder output a vector that could be any size.

Few-Shot Learning

In Machine Learning (ML), Few-Shot Learning is technique used to solve tasks with a very small amount of training data.

Generative

In the context of the Thesis, we are using the generic word Generative as the ability concept of an algorithm able generating outputs in a meaningful but unpredictable manner from an input, which includes Language Model (LM)s and Generative Models.

Generative Model

In Machine Learning (ML), Generative Models are generating random outputs from a single input by using the probability of observing the output based on the input. In other words, it models the probability of observation for a given target.

Glossary

Ground Learning

In the context of the AI, Grounded Learning is based on the Grounded theory from the social sciences, which uses inductive reasoning. In the context of AI, it is the mechanism of combining structured and unstructured data as small conceptual parts to then apply machine reasoning.

Hop

In Question Answering (QA) Systems, a Hop is a quantitative measure of the number of combinations necessary between indirectly related pieces of information to provide an answer.

Knowledge Base

In Information Systems (IS), a Knowledge Base is a Knowledge Representation using a Linked Data database for storing and interlinking structured and unstructured data using a standard.

Knowledge Graph

In Information Systems (IS), a Knowledge Graph is a Knowledge Base (KB) organized as a graph using semantics.

Language Model

In Natural Language Processing (NLP), a Language Model is a Model trained to provide likelihood probabilities of the following sequence of words in addition at providing the probability for each sequences of words.

Linked Data

In Information Systems (IS), Linked Data is a structured interlinked database mainly used for semantic queries.

Machine Reasoning

In ML, Machine Reasoning represent the ability to apply reasoning for a given input by using knowledge representations and logic patterns such as inductions, analogies, or abductions.

Machine Understanding

In ML, Machine Understanding stays ambiguous in its definition. However, we use the term as the ability of representing knowledge at an atomic building blocks and fundamental relations.

Markov Decision Process

In the context of the Reinforcement Learning (RL), this process models the ability of a predicting the next state of a finite-state machine-like process, such as a game, with only the information contained in the present state.

Model

In Machine Learning (ML), a model is the representation of the assumptions made by the algorithm during the training phase. Models are used to output a result based on a provided input and the learned patterns.

Model Fine-Tuning

In Machine Learning (ML), Fine Tuning a Model is the technique of using a trained Neural Network (NN) model as a base and tune it for a specific task.

Multi-Hop

In Question Answering (QA) Systems, a Multi-Hop implies that the answer is within multiple Hop of the question. In other words, the answer requires a combination of different information to be answerable. Generally, extra qualifying Subject-Predicate-Object Tuple (SPO) are separating the question Subject and the answer Object.

Named-Entity Linking

In Natural Language Processing (NLP), Name-Entity Linking extends the Named-Entity Recognition by providing an unique identifier to each word allowing a mapping in various databases (useful in translations).

Named-Entity Recognition

In Information Extraction (IE), Named-Entity Recognition is a technique used to extract from unstructured text words predefined in a vocabulary.

Open Domain

In Information Retrieval (IR), the support of Open Domain questions provides a no restrictions for the theme of the question asked.

Part of Speech

In Natural Language Processing (NLP), Part of Speech is a technique used to categorize words that behave syntactically similarly.

Part of Speech Tagging

In Natural Language Processing (NLP), The Part of Speech Tagging is extending the Part of Speech by adding a label to the word depending on its context (the neighboring words).

Reinforcement Learning

In ML, this type of learning combines generally a Markov Decision Process (MDP) environment with an approach similar to Unsupervised Learning (UL) as it does not require labelled data. The particularity of this technique is that it uses a notion of rewards to predict the best next-step by running a large amount of simulation as training.

Sequence-to-Sequence

In Machine Learning (ML), a Sequence-to-Sequence or Seq2Seq is an Encoder-Decoder Neural Network (NN) that for a given sequence of elements as input, outputs another sequence of elements.

Shallow Neural Network

In ML, similar to Deep Learning (DL), Shallow Neural Networks have a Encoder-Decoder approach by having a single hidden layer, which often has a high amount of parameters.

Glossary

Single-Hop

In Question Answering (QA) Systems, a Single-Hop implies that the answer is within a single Hop of the question. Generally, a unique Predicate separates the question Subject and the answer Object.

Supervised Learning

In ML, this type of learning implies the uses of labelled datasets to perform the training.

Transformer

In Natural Language Processing (NLP), a Transformer is a Sequence-to-Sequence (Seq2Seq) architecture using the Attention Mechanism..

Unsupervised Learning

In ML, this type of learning implies the uses of unlabelled datasets to perform the training.

Word Embedding

In Natural Language Processing (NLP), the Word Embedding is a technique for word representation as vectors in an embedding matrix. Additionally, it has often the particularity of preserving the semantical analogies of word-vectors.

Zero-Shot Learning

In Machine Learning (ML), Zero-Shot Learning is technique used to solve tasks without training on examples.

Acronyms

AGI
Artificial General Intelligence.

AI
Artificial Intelligence.

AIML
Artificial Intelligence Markup Language.

AL
Adversarial Learning.

ANI
Artificial Narrow Intelligence.

ANN
Artificial Neural Networks.

BERT
Bidirectional Encoder Representations from Transformers.

CBOW
Continuous Bag of words.

CE
Character Embedding.

CNN
Convolutional Neural Network.

DL
Deep Learning.

DNN
Deep Neural Networks.

FAQ
Frequently Asked Questions.

Acronyms

GAN
Generative Adversarial Networks.

GPT-2
Generative Pre-Training 2.

ICT
Information and Communications Technologies.

IE
Information Extraction.

IR
Information Retrieval.

IS
Information Systems.

KB
Knowledge Base.

KG
Knowledge Graph.

LM
Language Model.

MDP
Markov Decision Process.

ML
Machine Learning.

MN
Memory Network.

MR
Machine Reasonsing.

MRR
Mean Reciprocal Rank.

MRU
Master Research Units.

MT
Master's Thesis.

MU
Machine Understanding.

NL
Natural Language.

NLP
Natural Language Processing.

NLU
Natural Language Understanding.

NN
Neural Network.

OOV
Out-of-Vocabulary.

POC
Proof of Concept.

QA
Question Answering.

RL
Reinforcement Learning.

RNN
Recurrent Neural Network.

Seq2Seq
Sequence-to-Sequence.

SL
Supervised Learning.

SNN
Shallow Neural Network.

SOTA
State of the Art.

SPO
Subject-Predicate-Object Tuple.

TF-IDF
Term Frequency-Inverse Document Frequency.

Acronyms

UL

Unsupervised Learning.

Weak AI

Weak Artificial Intelligence.