Neural Analogy Generation: Phase 1

Ishan Singh Tezuesh Varshney

Prof. M.M. Sufyan Beg Mr. Misbahul Haque

Zakir Husain College of Engineering and Technology
Department of Computer Engineering
Aligarh Muslim University

November 23, 2019

Overview

Motivation and Objective

Despite the importance of storytelling as part of the human experience, computers still cannot create and tell novel stories, nor understand stories told by humans.

Our objective is to make computers better communicator, educator, entertainers and more capable of relating to us by genuinely understanding our needs. Thus we try to instill computers with *narrative intelligence* — the ability to craft, tell, and understand analogies.

Introduction : Breaking down the problem

The problem lies at the intersection of :

Natural Language Processing

- To understand and process human languages.

Cognitive Science

- To understand analogies.

Deep Learning and Natural Language Processing

- Use of Deep Neural Networks (DNNs) has changed the starting point for NLP problems a bit
  - Convert sparse representations to dense continuous ones
- Often use a pre-training technique like word2vec to create a distributed representation and plug those in.

Downstream Tasks solved by NLP

Deep Learning success in NLP - A non-exhaustive list:

- Part-of-speech Tagging
- Machine Translation
- Document Classification
- Question Answering etc.

Word2Vec Objective

- **Continous Bag of Words (CBOW)**: Given fixed surrounding window context, predict the middle word.[Mikolov et. a
- **Skip-gram**: Given middle word, predict fixed surrounding window.[Mikolov et. al.]

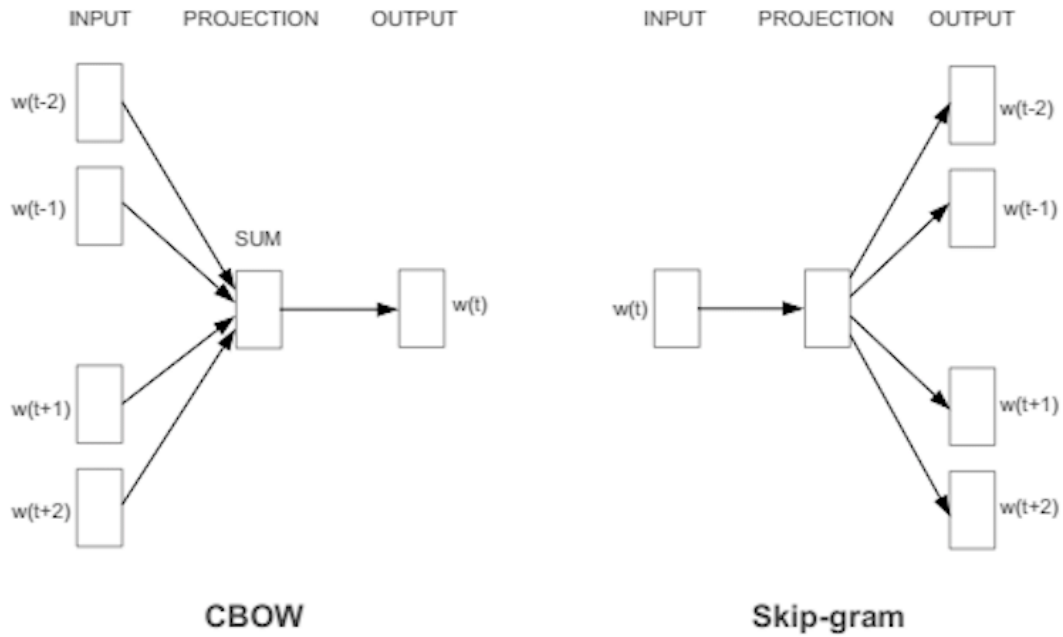


Figure: Representation of Words

One Hot Vectors

- One-hot: with $|V|$ array of vocabulary, only one "on" (1), the rest "off" (0)
- Represents the word at the temporal position t in T
- $|T|x|V|$ array representing a sentence

$$\begin{pmatrix} the \\ cat \\ sat \\ on \\ the \\ mat \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix}$$

Figure: One Hot Embedding

Lookup table-based Word Embeddings

- One-hot vector multiply by weight matrix yields row.
- Equivalent to looking up by the index.
  - Efficient, tensor contains only indices for "on" values.

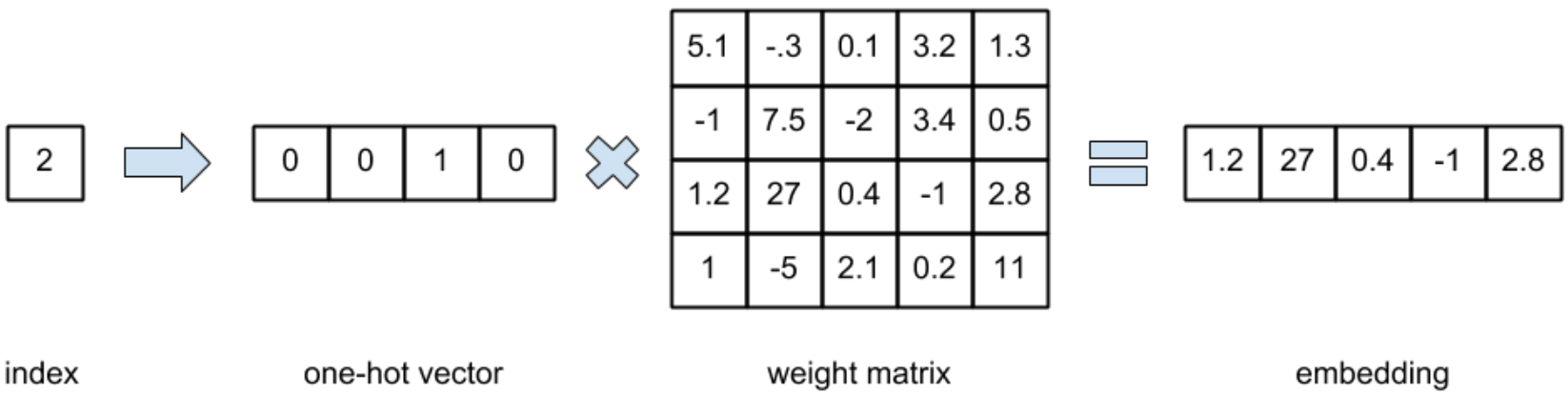


Figure: Look-Up

Language Modeling

The art to determining probability of next token.

Long Short Term Memory (LSTMs)

LSTMs Networks become a popular neural network architecture to learn this probabilities.

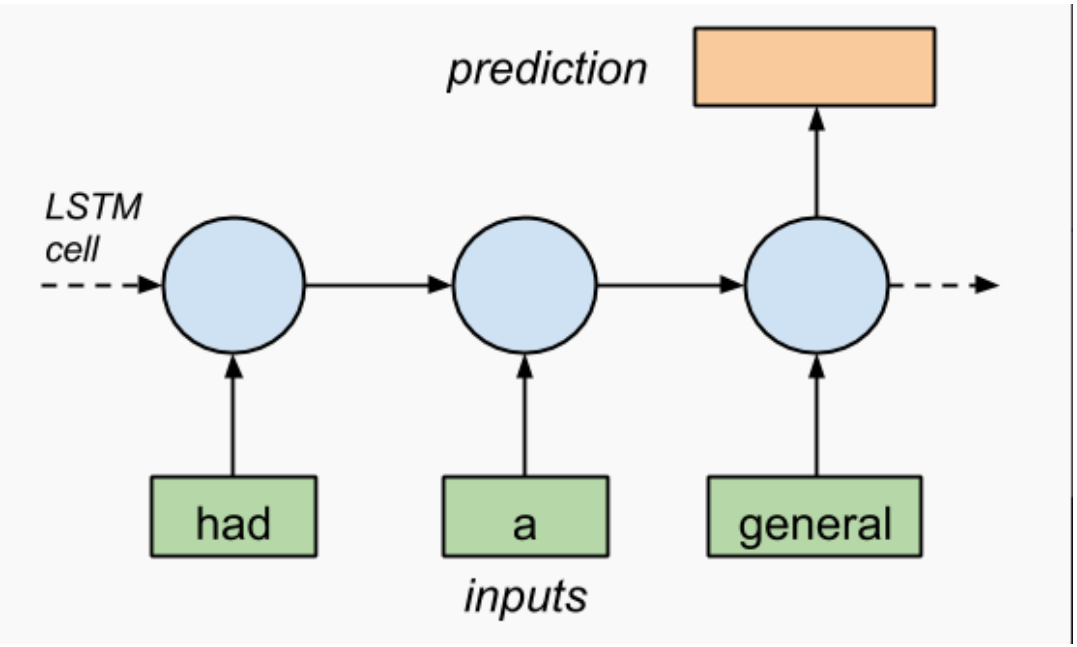


Figure: LSTM to predict next word

Machine Translation

Sequence-to-Sequence Network (seq2seq)

A seq2seq model consist of an Encoder-Decoder network

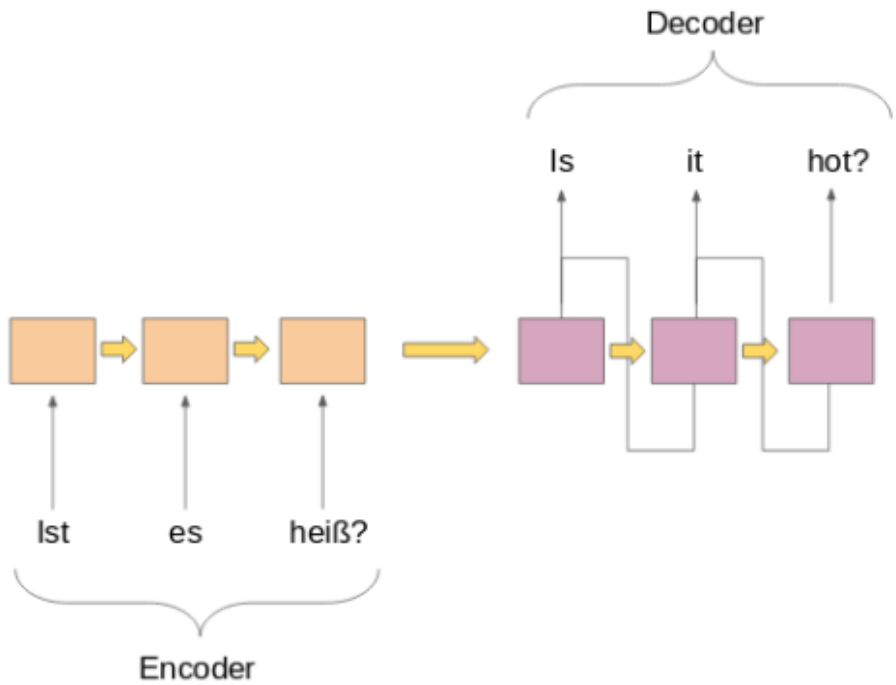


Figure: Encoder-Decoder Network for Machine Translation

Attention

- Attention[Bahdanau et. al.] is a mechanism that was developed to improve the performance of the Encoder-Decoder RNN on machine translation.
  - **Limitation:** Fixed length encoding make it difficult for network to cope with long sentences.

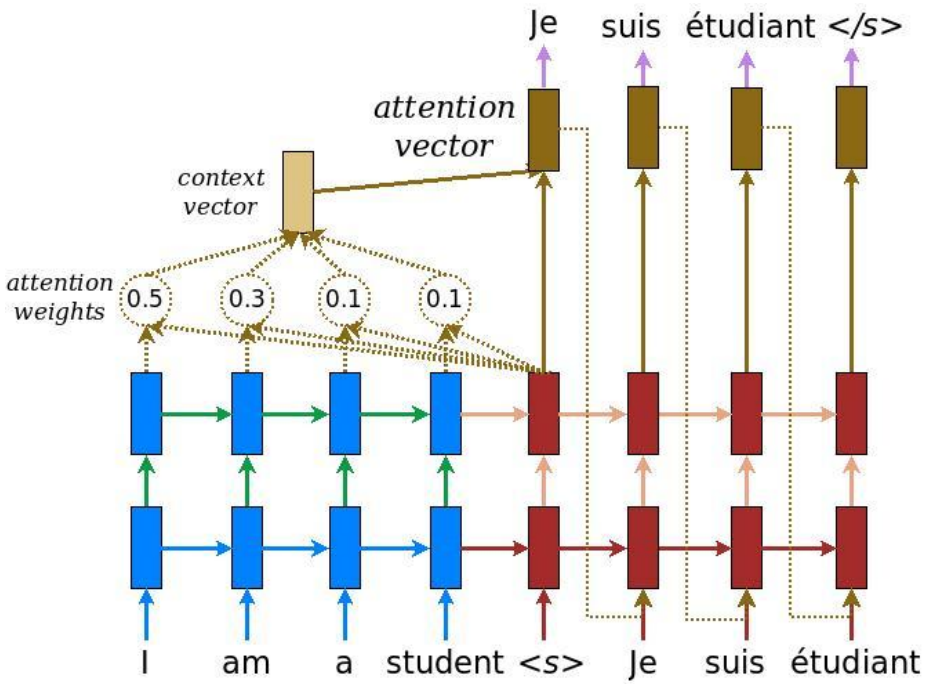


Figure: Attention based Encoder-Decoder Net. for Machine Translation

Attention is All You Need[Vaswani et. al.]

- **Goal:** To eliminate the LSTMs
  - Hard to parallelize due to autoregressive nature
  - Even with LSTM, long distance dependencies are challenging
- **Self-Attention:** is the method to bake the "understanding" of other relevant words into the one we're currently processing.

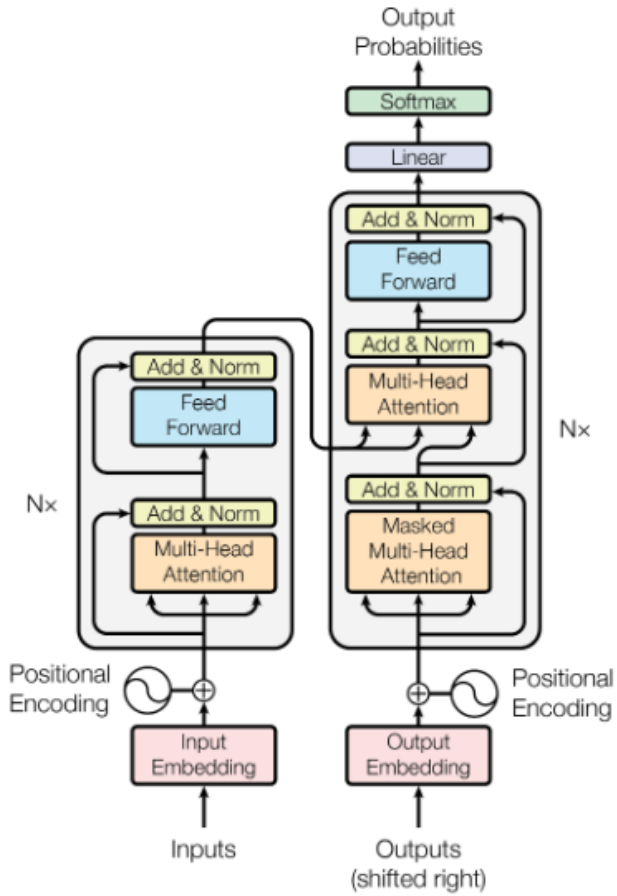


Figure: Architecture[Vaswani et. al.]

Deep Contextualized Language Model (LM)

Universal Language Model Fine-tuning for Text Classification (ULMFiT)[Howard et. al.]

The idea is to use generative pretrained LM + task-specific fine-tuning.
Base Model is ASGD Weight-Dropped LSTM (AWD-LSTM)[Socher et. al.].

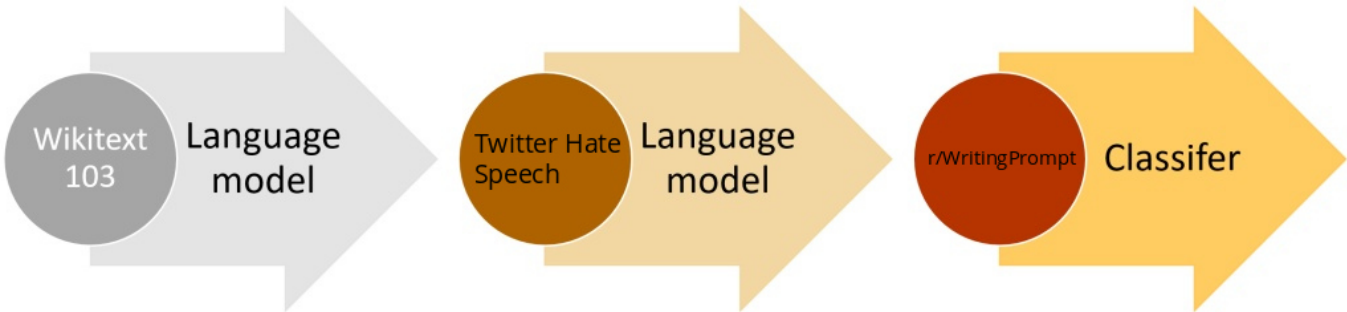


Figure: Fine-tuning ULMFiT

Deep Contextualized Language Model (LM)

Generative Pre-Training (GPT-2)[OpenAI]

It is a large transformer-based language model trained on 40 GBs of data

Parameters	Layers	d_{model}
117 M (Small)	12	768
345 M (Medium)	24	1024
774 M (Large)	36	1280
1.5 B (XL)	48	1600

Table: Variants of GPT2

Fine Tuning GPT-2

Abstract Generation

```
[ ] gpt2.download_gpt2(model_name="124M")

Fetching checkpoint: 1.05Mit [00:00, 576Mit/s]
Fetching encoder.json: 1.05Mit [00:00, 136Mit/s]
Fetching hparams.json: 1.05Mit [00:00, 575Mit/s]
Fetching model.ckpt.data-00000-of-00001: 498Mit [00:02, 194Mit/s]
Fetching model.ckpt.index: 1.05Mit [00:00, 385Mit/s]
Fetching model.ckpt.meta: 1.05Mit [00:00, 166Mit/s]
Fetching vocab.bpe: 1.05Mit [00:00, 223Mit/s]
```

```
[ ] sess = gpt2.start_tf_sess()
```

```
[ ] gpt2.finetune(sess,
                  dataset='drive/My Drive/FineTuneAbstract/arXiv.txt',
                  model_name='124M',
                  steps=500,
                  restore_from='fresh',
                  run_name='run1',
                  sample_every=100,
                  )
```

Figure: Model and Training process

```
[ ] pref = "<|startText|>BinGAN: Learning Compact Binary Descriptors with a \
           Regularized GAN\n"
```

```
stri = gpt2.generate(sess,
                     checkpoint_dir = 'drive/My Drive/FineTuneAbstract/checkpoint',
                     run_name='run1',
                     prefix=pref,
                     return_as_list=True)[0]
```

```
[31] textwrap.fill(stri[13:], width=80)

BinGAN: Learning Compact Binary Descriptors with a Regularized GAN  This paper
presents a novel approach to learning compact binary-encodings of the original
image. We developed a method to incorporate a recently proposed technique for
learned binary representations in the modern image processing pipeline.
Specifically, we build a neural network with a constructed binary representation
from an image origin and a previously designed binary representation from a
different image origin. The proposed approach is adapted to the recurrent neural
network on the Convolutional Neural Network (CNN) and the GAN. Extensive
experiments on different image datasets and problem domains demonstrate that the
proposed approach accurately learns compact binary representations. Furthermore,
the extracted binary representation is non-destructive and can be used for
training-to-region (BFR) algorithms. The proposed approach is applied to the CNN
Markov chain Monte Carlo (MCMC) and the GAN datasets. The learned binary
representations are compared with previous Convolutional BFR learning methods to
identify the best performing BFR image of the original image. Experimental
results on two image datasets demonstrate the superiority of our approach over
the state-of-the-art methods.
```

Figure: Abstract Generation

Limitations of Fine-tuning large pre-trained network

- Their still exist a large gap between performance of baseline method and human performance
- Reasoning with neural network lack transparency and interpretability

Knowledge Aware Model

To enhance the reasoning capacity, these networks are used to explicitly incorporate external knowledge as relational *inductive biases*. [Battaglia et. al.]
Furthermore, a knowledge-centric approach is extensible through knowledge acquisition techniques[Xiang Li].

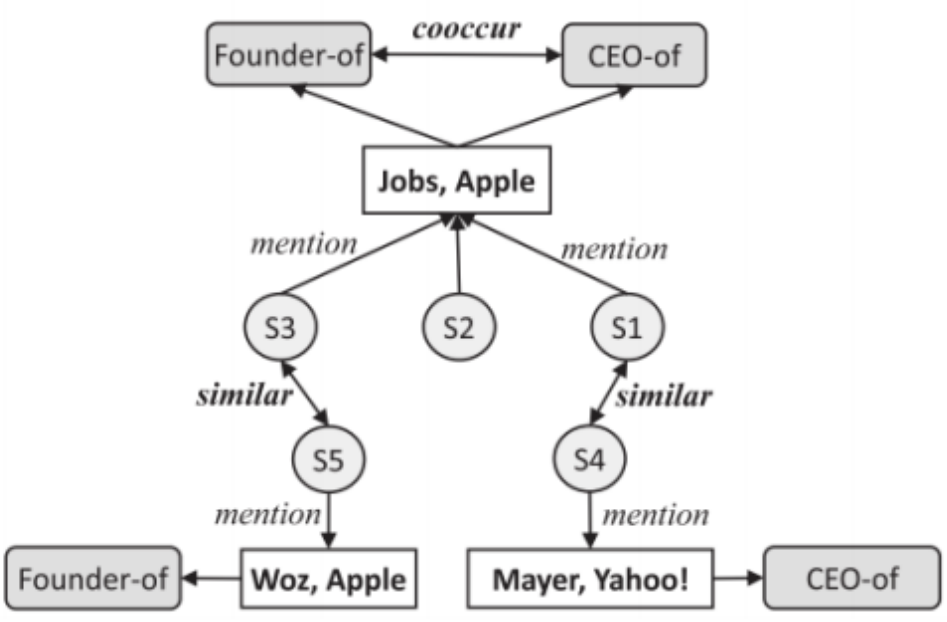


Figure: Dependencies between objects in Knowledge Base

Analogy

- The ability to make analogies – that is, to flexibly map familiar relations from one domain of experience to another – is a fundamental ingredient of human intelligence and creativity
- The cognitive process of Analogy can be explained as:
  - **Structure Mapping Theory (SMT):** It emphasizes the distinction between two means of comparing domains of experience; analogy and similarity.[Genter (1983)][10]. It seeks a more "horizontal" view[12].
  - **High-Level Perception Theory (HLP)**: the process of making sense of complex data at an abstract, conceptual level—is fundamental to human cognition[11][10]. It seeks a more "vertical" view[12]

Analogy and Knowledge Graph

Analogy being a linear mapping can be expressed a knowledge graph.

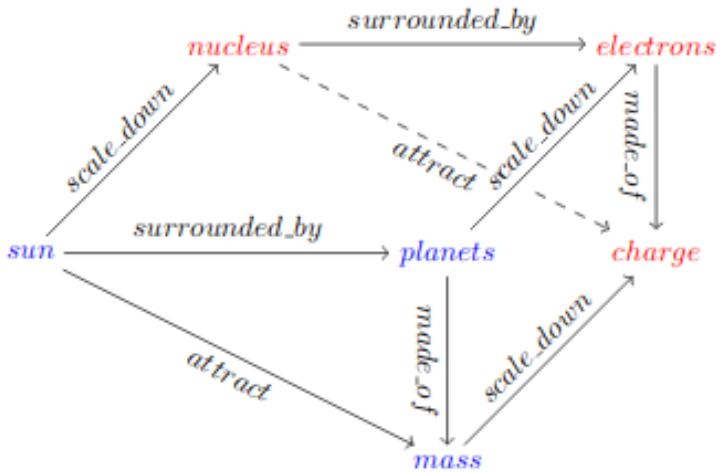
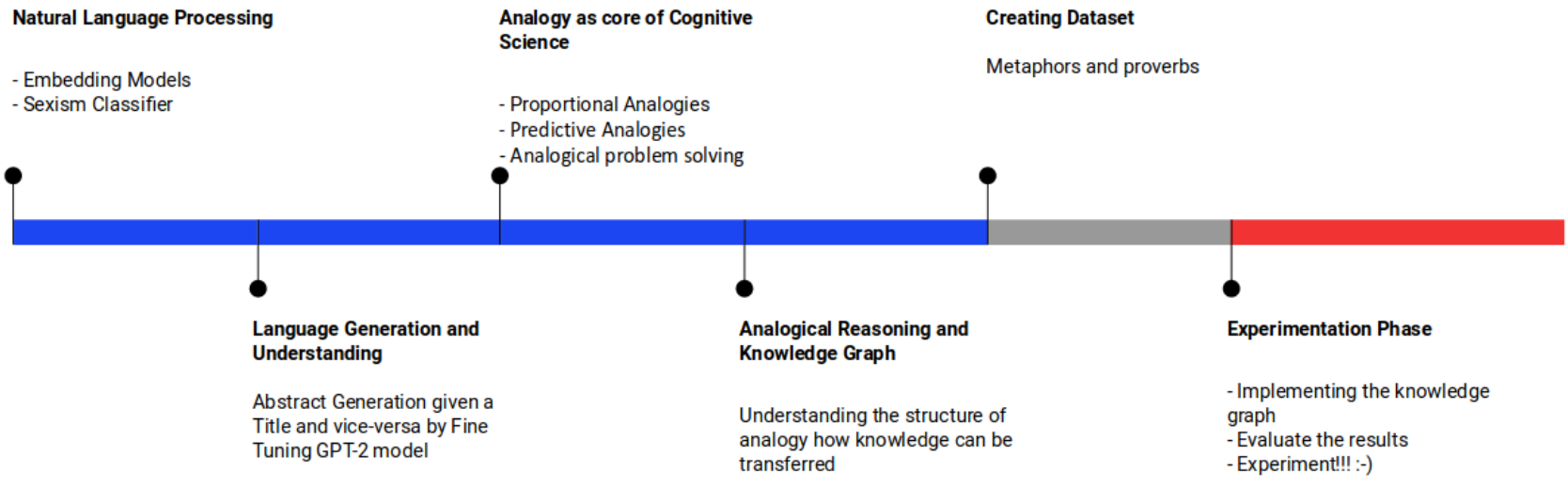


Figure: Commutative diagram for the analogy between the Solar System (red) and the Rutherford-Bohr Model (blue)

Underlying Principle

- To develop a dependency graph between the different part of the sentence.
- Learn the representation of the graph for interpretability.
- Use the learn't representation to develop a language model for analogical reasoning.

Timeline


Natural Language Processing
- Embedding Models
- Sexism Classifier
Analogy as core of Cognitive Science
- Proportional Analogies
- Predictive Analogies
- Analogical problem solving
Creating Dataset
Metaphors and proverbs
Language Generation and Understanding
Abstract Generation given a Title and vice-versa by Fine Tuning GPT-2 model
Analogical Reasoning and Knowledge Graph
Understanding the structure of analogy how knowledge can be transferred
Experimentation Phase
- Implementing the knowledge graph
- Evaluate the results
- Experiment!!! :-)

References


Tomas Mikolov, Kai Chen, Gregory S. Corrado and Jeffrey Dean (2013)
Efficient Estimation of Word Representations in Vector Space
CoRR


Dzmitry Bahdanau, Kyunghyun Cho and Yoshua Bengio (2015)
Neural Machine Translation by Jointly Learning to Align and Translate
International Conference on Learning Representations (ICLR)


Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Lukasz Kaiser and Illia Polosukhin (2017)
Attention Is All You Need
Conference on Neural Information Processing Systems (NIPS


Jeremy Howard and Sebastian Ruder (2018)
Universal Language Model Fine-tuning for Text Classification
Annual Conference of the Association for Computational Linguistics


Stephen Merity, Nitish Shirish Keskar and Richard Socher (2018)
Regularizing and Optimizing LSTM Language Models
International Conference on Learning Representations

References

Alec Radford, Jeffrey Wu, Rewon Child, David Luan, Dario Amodei and 1Ilya Sutskever (2018)
Language Models are Unsupervised Multitask Learners

Xiang Li, Aynaz Taheri, Lifu Tu, and Kevin Gimpel (2016)
Commonsense knowledge base completion.
Annual Conference of the Association for Computational Linguistics

Peter W. Battaglia, Jessica B. Hamrick, Victor Bapst, Alvaro Sanchez-Gonzalez, Vincius Flores Zambaldi, Mateusz Malinowski, Andrea Tacchetti, David Raposo, Adam Santoro, Ryan Faulkner, C¸ aglar Gulc¸ehre, Francis Song, Andrew J. Ballard, Justin Gilmer, George E. Dahl, Ashish Vaswani, Kelsey R. Allen, Charles Nash, Victoria Langston, Chris Dyer, Nicolas Heess, Daan Wierstra, Pushmeet Kohli, Matthew Botvinick, Oriol Vinyals, Yujia Li, and Razvan Pascanu. (2018)
Relational inductive biases, deep learning, and graph networks.
CoRR

Dedre Gentner (1983)
Structure-mapping: A theoretical framework for analogy.
Cognitive science, 7(2):155–170

References

Felix Hill, Adam Santoro, David G.T. Barrett, Ari S. Morcos and Timothy Lillicrap (2019)
Learning to Make Analogies by Contrasting Abstract Relational Structure
International Conference on Learning Representations

David J Chalmers, Robert M French, and Douglas R Hofstadter (1992)
High-level perception, representation, and analogy: A critique of artificial intelligence methodology.
Journal of Experimental Theoretical Artificial Intelligence, 4(3):185–211

Clayton T. Morrison and Eric Dietrich (1995)
Structure-Mapping vs. High-level Perception:The Mistaken Fight Over The Explanation of Analogy
Annual Conference of the Cognitive Science Society, pp.678-682

The End