UNIVERSITY of DAYTON

VISION LAB
Center of Excellence

Brad Sorg
Dr. Vijayan K. Asari

Natural Language Processing: A Look Into How Computers Understand Human Language

Overview

- Subfield of machine learning concerned with interactions between computers and human (natural) language
- Program computers to process & analyze large amounts of natural language data
- Importance
  - Large volumes of textual data – analyze more language based data than humans
  - Structuring of highly unstructured data sources
- Breaks down language into shorter, elemental pieces
  - Understands relationships between pieces – explores how pieces work together to create meaning
- Counts, groups, & categorizes words to extract structure/meaning from large volumes of content
- Goes beyond structural understanding of language
  - Interprets intent, resolves context & word ambiguity
  - Generates well-formed human language on its own

Applications

- Text Classification
  - Spam filtering, categorizing news articles, etc.
- Language Modeling
  - Text recognition, generate suggested continuation of a sentence, etc.
- Speech Recognition
  - Transcribing a speech, creating text for TV show/movie
- Machine Translation
  - Translating text/audio from one language to another
- Question Answering
  - Given a subject, answer a specific question about subject (Google)

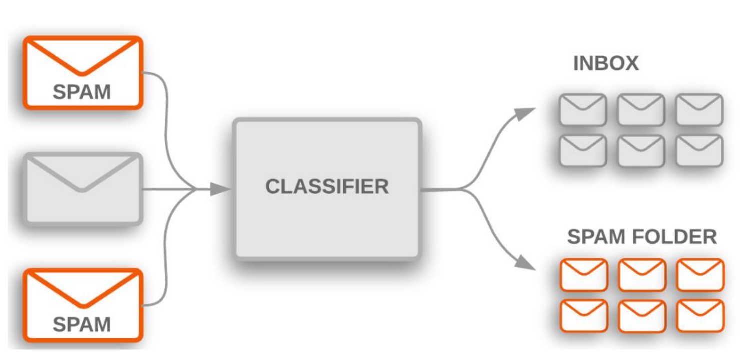

Fig 1: Spam Filtering Example

Sequence to Sequence (Seq2Seq) Learning

- Training models to convert sequences from one domain to sequence in another
  - Ex. Translating sentence from English to French, creating subtitles for movie
- Free-form question answering systems
  - Generate answer when given a question
- Applicable any time text needs to be generated
- Implemented using Recurrent Neural Networks (RNNs)

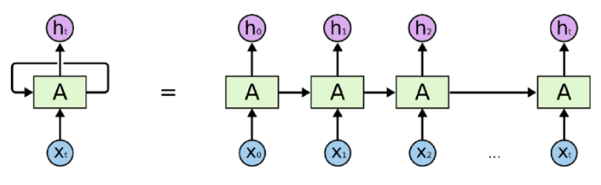

Fig 2: RNN Architecture

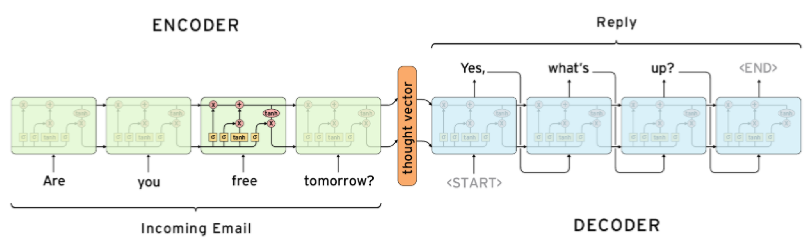

Fig 3: Seq2Seq Example

Temporal Convolution Network (TCN)

- Generic architecture for convolutional sequence prediction
- Distinguishing characteristics:
  - Causal convolutions – no information "leakage" from future into the past
  - Can take a sequence of any length and map it to an output sequence of the same length
- Causal convolution – convolutions where an output at time *t* is convolved only with elements from time *t* and earlier in the previous layer

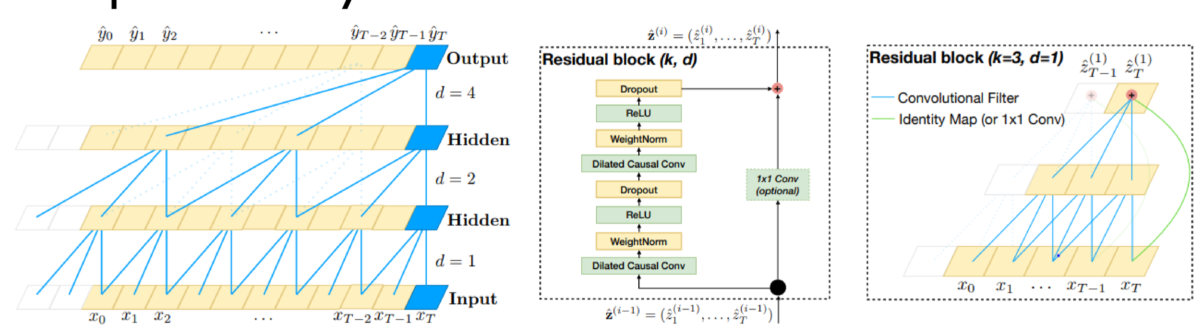

Fig 4: TCN Functionality

Embeddings from Language Models (ELMo)

- Deep contextualized word representation that models both:
  - Complex characteristics of word use – syntax/semantics
  - How these uses vary across linguistic contexts
- Use bidirectional language model (biLM) to learn both word and linguistic context
- Instead of using a fixed embedding for each word, ELMo looks at the entire sentence before assigning each word in it an embedding

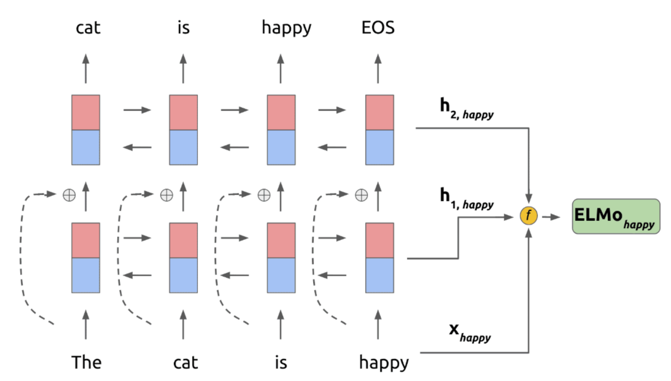

Fig 5: Sample ELMo Task

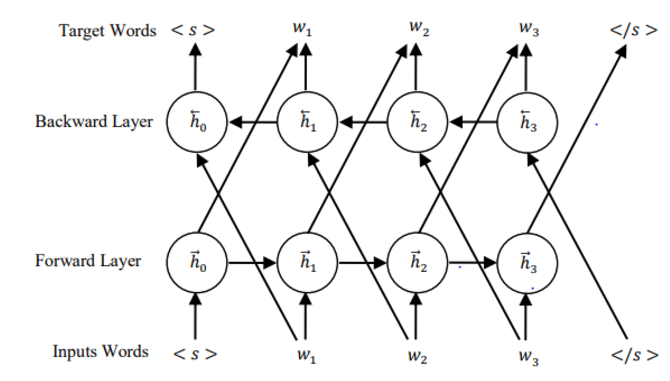

Fig 6: biLM Architecture

BERT

- Bidirectional encoder representation from transformers (BERT)
- A multi-layer bidirectional transformer encoder
- Designed to pre-train deep bidirectional representations by jointly conditioning on both left & right context in all layers
- Transformer encoder reads the entire sequence of words at once
  - Allows model to learn context of a word based on all of its surroundings

Input: The man went to the [MASK]$_1$. He bought a [MASK]$_2$ of milk .
Labels: [MASK]$_1$ = store; [MASK]$_2$ = gallon

Sentence A = The man went to the store.
Sentence B = He bought a gallon of milk.
Label = IsNextSentence

Sentence A = The man went to the store.
Sentence B = Penguins are flightless.
Label = NotNextSentence

Fig 7: Sample BERT Tasks

UNIVERSITY of DAYTON

Vision Lab
Kettering Laboratories
300 College Park
Dayton, OH 45469-0232

Center of Excellence for Computer Vision and Wide Area Surveillance Research
visionlab.udayton.edu

Dr. Vijayan K. Asari
Director
(937) 229-4504
vasari1@udayton.edu