Deep Learning for Joint-Source Channel Coding of Text

Nariman Farsad*, Milind Rao*, Andrea Goldsmith
{nfarsad, milind, andreag}@stanford.edu

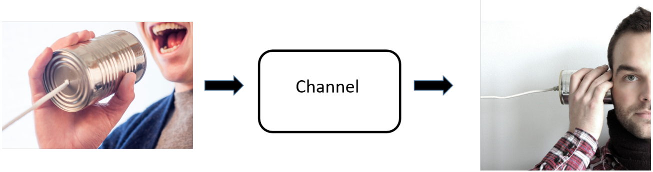
Wireless Systems Lab
Stanford University

International Conference on Acoustics, Sound, and Signal Processing
Calgary, 2018

Outline

1. Problem Description
   - Model
   - Performance Metrics
2. Deep Encoder and Decoder
3. Results
   - Baselines
   - Properties of the Encoding

Transmission of Text



Good channel/no rate constraint

'Advanced avians, ambulating in the ante meridiem, are advantaged in apprehending an annelid' → 'Advanced avians, ambulating in the ante meridiem, are advantaged in apprehending an annelid'

Noisy channel/rate constrained

'Advanced avians, ambulating in the ante meridiem, are advantaged in apprehending an annelid' → 'The early bird catches the worm'

Model Description

Data

- Vocabulary $\mathcal{V} = \{\text{'.'}, \text{'94'}, \text{'the'}, \text{'european'}, \ldots\}$
- Sentence $\mathbf{s} = [w_1, w_2, \ldots, w_n],\ w_i \in \mathcal{V}$ is to be transmitted

Encoder

- Encoder $\psi_\ell : \mathcal{V}^* \to \{0, 1\}^\ell$
- takes variable length sentence $\mathbf{s}$
- produces $\ell-$length binary encoding, $\mathbf{b} = \psi_\ell(\mathbf{s})$

Model Description Continued

Channel

Erasure channel with erasure probability ρ

$$\mathbf{o}_i = \begin{cases} \mathbf{b}_i & \text{w.p. } 1-\rho \\ \text{err} & \text{w.p. } \rho \end{cases} \text{ for } i \in \{1, \ldots, \ell\}$$

Decoder

- Decoder $\nu_\ell : \{0, 1, \text{err}\}^\ell \to \mathcal{V}^*$
- takes channel output $\mathbf{o}$
- produces sentence $\hat{\mathbf{s}} = \nu_\ell(\mathbf{o}) = [\hat{w}_1, \ldots, \hat{w}_{\hat{n}}]$ of possibly different length

Performance Metrics

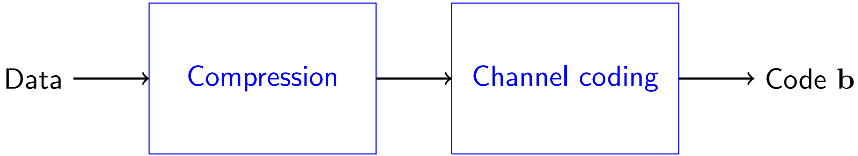
Word Accuracy

- loss $= \sum_{i=1}^{n} \mathbf{1}(w_i \neq \hat{w}_i)$
- 'An example sentence for you', 'This is an example sentence for you'. Loss $= 5$!

Edit Distance

- Minimum length of sequence of insert, delete, replace operations to transform $\mathbf{s} \to \hat{\mathbf{s}}$
- 'An example sentence for you', 'This is an example sentence for you'. Loss $= 2$
- Does not capture effect of synonyms

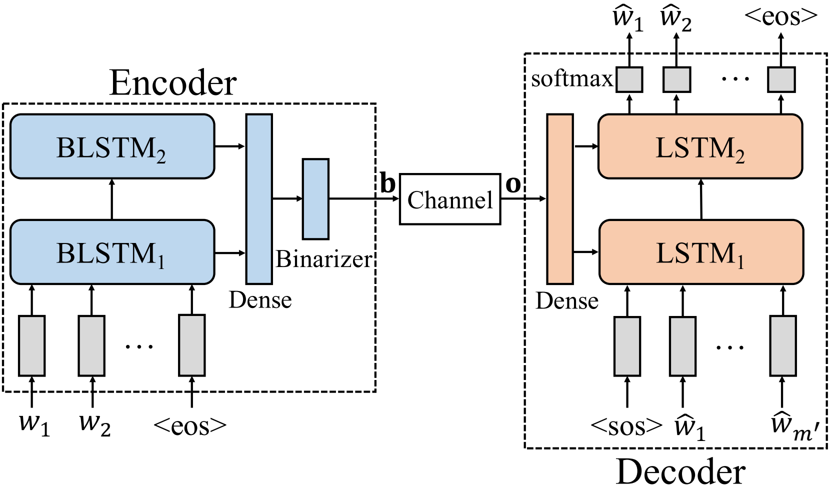
Joint vs Separate Source-Channel Coding of Text



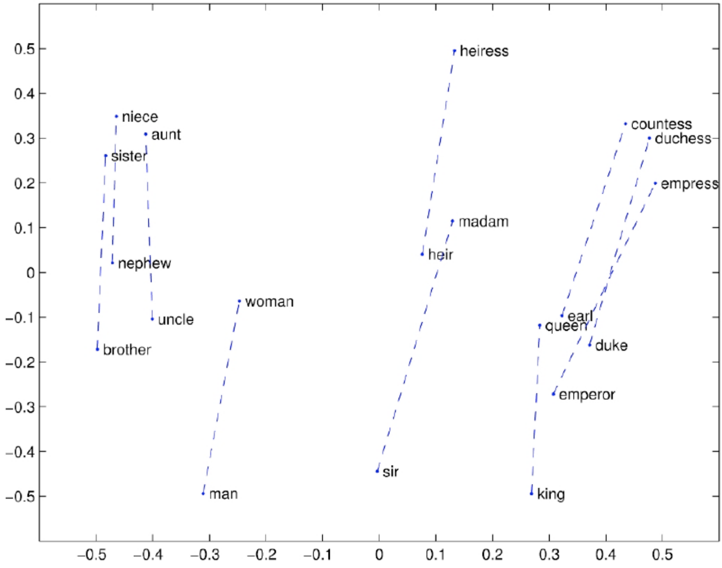
- Shannon: Separate source-channel coding is optimal
  - For some (eg. ergodic and memoryless) channels
  - Infinite block lengths and delay
  - No limit on code complexity
- We propose a deep neural network for joint SC coding
  - Goal: Convey semantic information of a sentence
  - Deep NLP - Neural networks capture complicated language probability models
  - Contrast with prior deep NLP (eg. Google translate) - focus on compression

Outline

Deep Learning Architecture



Building Block: Glove Word Embeddings



- 200-dimensional vectors represent meaning of words
- Need to combine word vectors to form a sentence vector

Deep Learning Architecture

1. Encoder
   - Robust sequential autoencoder
   - Sentences of any length are mapped to a binary encoding of fixed length
2. Channel
   - Channel implemented using dropout
   - Can be expanded to other channels - AWGN, binary symmetric channel, etc
3. Decoder
   - At each point, decoder outputs logits/probability of words $p(w)$
   - Cross-entropy loss

$$\mathsf{Loss}_i = \sum_{w \in \mathcal{V}} -\mathbf{1}(w_i = w) \log p(w)$$

   - Performance improved by using beam decoder

Outline

1. Problem Description
   - Model
   - Performance Metrics
2. Deep Encoder and Decoder
3. Results
   - Baselines
   - Properties of the Encoding

Baselines for Comparison

Source-Coding or compression:

- Universal compressor (gzip):
  - Reaches entropic limit in the asymptote
  - Needs large (30+) batches of sentences, not single one
- Huffman coding
- 5 bit encoding for characters

Channel-coding through Reed-Solomon codes.

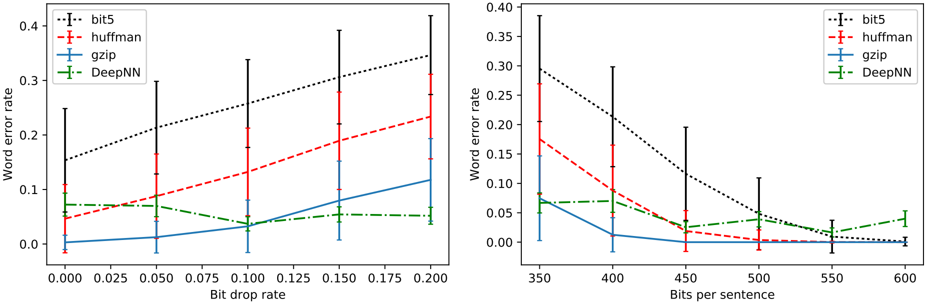
Errors

- If number of bits/sentence is low: part of sentence cannot be transmitted
- Channel decoding error: whole sentence lost for Huffman, universal coding

Examples of Deep Joint SC Errors

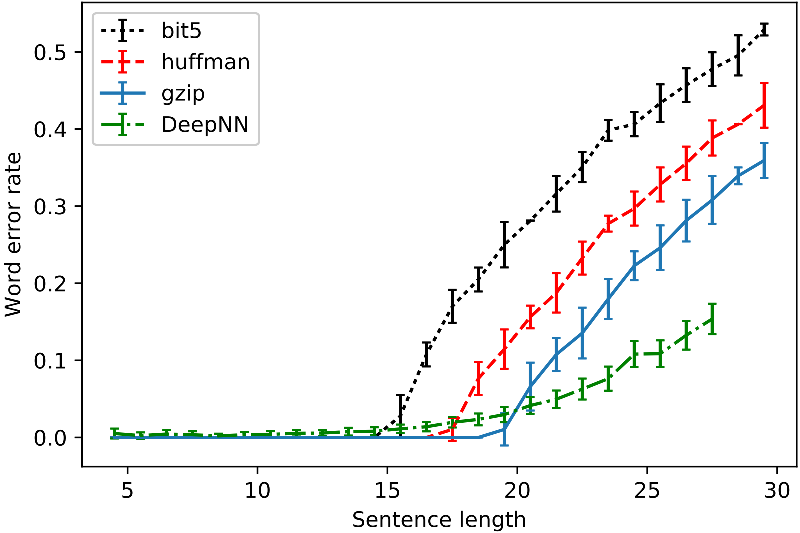
Punctuation error	TX: efficiency what efficiency ? RX: efficiency , what efficiency ?
Rephrasing	TX: tourism serves as a source of income to totalitarian regimes . RX: tourism has become a source of income to totalitarian regimes .
Rephrasing	TX: a few wealthy individuals compared with millions living in hunger . RX: a few wealthy individuals face with millions living in hunger .
Tense Error	TX: a communist country riding roughshod over human rights . RX: a communist country rides roughshod over human rights .
An inexplicable error	TX: i listened to colleagues who mentioned bicycles . RX: i listened to colleagues who mentioned goebbels .
Long sentence	TX: there is one salient fact running through these data : the citizens want more information and have chosen television as the best means to receive that information . RX: there is one glaring weaknesses , by the communication : the citizens want more information and hold ' television as the means to receive this information .

Results

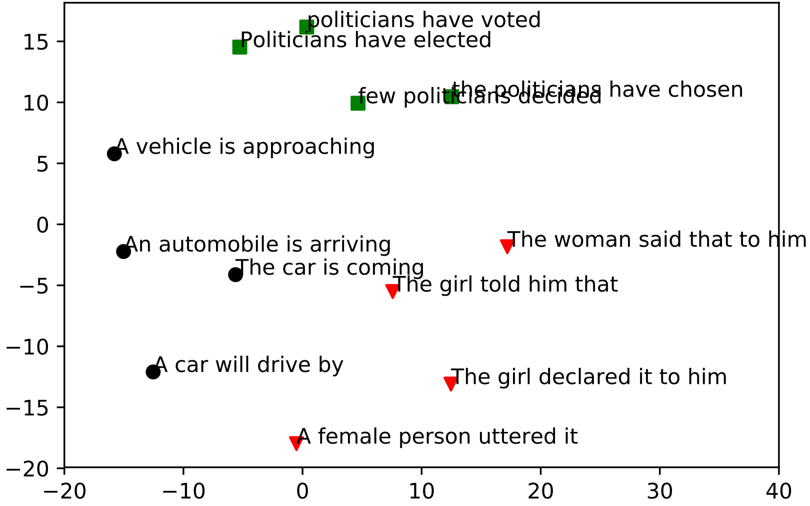


In very rate constrained regimes - deep NN outperforms baselines

Impact of Fixed Length Encoding



Properties of the Encoding



Summary and Future Work

- Proposed robust autoencoder based joint source-channel coding for text
- Encoding is done in a *sentence space*
- Recovery of information more important than exact sentence recovery
- Scheme outperforms baselines when number of bits per sentence is low

Future Work

- Rethink performance metrics
- Variable length encoding
- Other kinds of structured data: audio, speech, video