# Building End-To-End Dialogue Systems Using Generative Hierarchical Neural Network Models

**Bogdan-Adrian Muşat**
Xperi
bogdan.musat@xperi.com

## Abstract

This project is an attempt to implement the ideas and replicate the results of Serban et al. [2015]. The paper introduces a new idea for preserving the context of an open domain, conversational dialogue system based on large dialogue corpora using generative models. The main technical challenge was to accurately implement and train the network, as described in the paper, using the Keras framework. Although I successfully implemented the paper's novel proposal, I was only partially able to replicate the training results that the original authors report.

## 1  Introduction

The impressive achievements of Recurrent Neural Networks (RNNs) have revolutionized the field of natural language processing in recent years. Due to their improved performance on tasks such as sentiment analysis, language translation, language modeling, they have become a staple approach for the industry and research community. Cho et al. [2014] proposed a novel network architecture called RNN Encoder-Decoder that consists of two RNNs, later known as seq2seq (Sutskever et al. [2014]). As the name suggests, one RNN encodes a sequence of symbols into a fixed length vector representation, and the other decodes the representation into another sequence of symbols. This kind of architecture represented a real breakthrough for the field of machine translation, and a big step in the direction of automated chatting systems. What this model lacks when it comes to building such a system is the preservation of the conversational context. While in machine translation the actual translation depends only on the current utterance, in a real life conversation you have to keep track of previous information obtained along the way. Sordoni et al. [2015] proposed a model called Hierarchical Recurrent Encoder-Decoder (HRED), which builds on top of the seq2seq idea. It uses a horizontal stack of seq2seq models, binded by a contextual RNN. This allows the information to flow from the past, like in the case of an RNN model, where here each utterance is a different time step. The model architecture was extended to better suit the dialogue task. To carry out experiments, I used the MovieTriples dialogue dataset based on movie scripts, which was available upon request from the authors.

## 2  Model and Training

A representation of the HRED model architecture is given in Figure 1. A training sample, in this case, consists of a three-turn utterances coming from a dialog. The encoder is responsible of representing the input phrase as a fixed length vector. To better capture the full understanding of the utterance, the encoder uses a Bidirectional RNN (Schuster and Paliwal [1997]). The output is obtained as the concatenation of the forward and backward runs of the Bidirectional RNN. The context RNN (middle part) maintains a history of the whole dialogue, by taking into consideration all the previous utterances that appeared up until now. The decoder's responsibility is to reproduce as accurately as possible the reply attached to the encoder's input. At each time step of the decoder, the network is
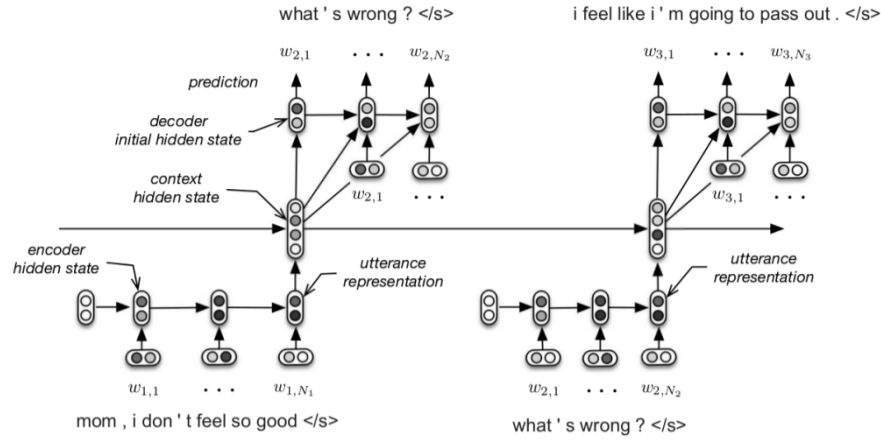
Figure 1: The computational graph of the HRED architecture for a dialogue composed of three turns.

trying to predict the next word in the utterance, using the current input word and the information from the context RNN. Thus, the decoder RNN models a conditional distribution based on a input $x_t$ and a fixed length vector $c$. For the embedding of the input words, the model uses the pretrained weights from word2vec Mikolov et al. [2013], to better capture the meaning of each particular word.

# 3 Results

# 4 Conclusion

# 5 Model and Training

# 6 Headings: first level

All headings should be lower case (except for first word and proper nouns), flush left, and bold.

First-level headings should be in 12-point type.

## 6.1 Headings: second level

Second-level headings should be in 10-point type.

### 6.1.1 Headings: third level

Third-level headings should be in 10-point type.

**Paragraphs** There is also a \paragraph command available, which sets the heading in bold, flush left, and inline with the text, with the heading followed by 1 em of space.

# 7 Citations, figures, tables, references

These instructions apply to everyone.

## 7.1 Citations within the text

The natbib package will be loaded for you by default. Citations may be author/year or numeric, as long as you maintain internal consistency. As to the format of the references themselves, any style is acceptable as long as it is used consistently.

The documentation for `natbib` may be found at

```
http://mirrors.ctan.org/macros/latex/contrib/natbib/natnotes.pdf
```

Of note is the command `\citet`, which produces citations appropriate for use in inline text. For example,

```
\citet{hasselmo} investigated\dots
```

produces

Hasselmo, et al. (1995) investigated...

If you wish to load the `natbib` package with options, you may add the following before loading the `nips_2017` package:

```
\PassOptionsToPackage{options}{natbib}
```

If `natbib` clashes with another package you load, you can add the optional argument `nonatbib` when loading the style file:

```
\usepackage[nonatbib]{nips_2017}
```

As submission is double blind, refer to your own published work in the third person. That is, use "In the previous work of Jones et al. [4]," not "In our previous work [4]." If you cite your other papers that are not widely available (e.g., a journal paper under review), use anonymous author names in the citation, e.g., an author of the form "A. Anonymous."

## 7.2 Footnotes

Footnotes should be used sparingly. If you do require a footnote, indicate footnotes with a number[1] in the text. Place the footnotes at the bottom of the page on which they appear. Precede the footnote with a horizontal rule of 2 inches (12 picas).

Note that footnotes are properly typeset *after* punctuation marks.[2]

## 7.3 Figures

All artwork must be neat, clean, and legible. Lines should be dark enough for purposes of reproduction. The figure number and caption always appear after the figure. Place one line space before the figure caption and one line space after the figure. The figure caption should be lower case (except for first word and proper nouns); figures are numbered consecutively.

You may use color figures. However, it is best for the figure captions and the paper body to be legible if the paper is printed in either black/white or in color.

## 7.4 Tables

All tables must be centered, neat, clean and legible. The table number and title always appear before the table. See Table 1.

Place one line space before the table title, one line space after the table title, and one line space after the table. The table title must be lower case (except for first word and proper nouns); tables are numbered consecutively.

Note that publication-quality tables *do not contain vertical rules.* We strongly suggest the use of the `booktabs` package, which allows for typesetting high-quality, professional tables:

```
https://www.ctan.org/pkg/booktabs
```

This package was used to typeset Table 1.

---

[1]Sample of the first footnote.
[2]As in this example.

Figure 2: Sample figure caption.

Table 1: Sample table title

| | Part | | Size ($\mu$m) |
|---|---|---|---|
| Name | Description | | Size ($\mu$m) |
| Dendrite | Input terminal | | $\sim$100 |
| Axon | Output terminal | | $\sim$10 |
| Soma | Cell body | | up to $10^6$ |

## 8 Final instructions

Do not change any aspects of the formatting parameters in the style files. In particular, do not modify the width or length of the rectangle the text should fit into, and do not change font sizes (except perhaps in the **References** section; see below). Please note that pages should be numbered.

## 9 Preparing PDF files

Please prepare submission files with paper size "US Letter," and not, for example, "A4."

Fonts were the main cause of problems in the past years. Your PDF file must only contain Type 1 or Embedded TrueType fonts. Here are a few instructions to achieve this.

- You should directly generate PDF files using `pdflatex`.
- You can check which fonts a PDF files uses. In Acrobat Reader, select the menu Files>Document Properties>Fonts and select Show All Fonts. You can also use the program `pdffonts` which comes with `xpdf` and is available out-of-the-box on most Linux machines.
- The IEEE has recommendations for generating PDF files whose fonts are also acceptable for NIPS. Please see `http://www.emfield.org/icuwb2010/downloads/IEEE-PDF-SpecV32.pdf`
- `xfig` "patterned" shapes are implemented with bitmap fonts. Use "solid" shapes instead.
- The `\bbold` package almost always uses bitmap fonts. You should use the equivalent AMS Fonts:

      \usepackage{amsfonts}

  followed by, e.g., \mathbb{R}, \mathbb{N}, or \mathbb{C} for $\mathbb{R}$, $\mathbb{N}$ or $\mathbb{C}$. You can also use the following workaround for reals, natural and complex:

      \newcommand{\RR}{I\!\!R} %real numbers
      \newcommand{\Nat}{I\!\!N} %natural numbers
      \newcommand{\CC}{I\!\!\!\!C} %complex numbers

  Note that `amsfonts` is automatically loaded by the `amssymb` package.

If your file contains type 3 fonts or non embedded TrueType fonts, we will ask you to fix it.

## 9.1 Margins in LaTeX

Most of the margin problems come from figures positioned by hand using `\special` or other commands. We suggest using the command `\includegraphics` from the `graphicx` package. Always specify the figure width as a multiple of the line width as in the example below:

```
\usepackage[pdftex]{graphicx} ...
\includegraphics[width=0.8\linewidth]{myfile.pdf}
```

See Section 4.4 in the graphics bundle documentation (http://mirrors.ctan.org/macros/latex/required/graphics/grfguide.pdf)

A number of width problems arise when LaTeX cannot properly hyphenate a line. Please give LaTeX hyphenation hints using the `\-` command when necessary.

### Acknowledgments

Use unnumbered third level headings for the acknowledgments. All acknowledgments go at the end of the paper. Do not include acknowledgments in the anonymized submission, only in the final paper.

# References

References follow the acknowledgments. Use unnumbered first-level heading for the references. Any choice of citation style is acceptable as long as you are consistent. It is permissible to reduce the font size to `small` (9 point) when listing the references. **Remember that you can go over 8 pages as long as the subsequent ones contain *only* cited references.**

[1] Alexander, J.A. & Mozer, M.C. (1995) Template-based algorithms for connectionist rule extraction. In G. Tesauro, D.S. Touretzky and T.K. Leen (eds.), *Advances in Neural Information Processing Systems 7*, pp. 609–616. Cambridge, MA: MIT Press.

[2] Bower, J.M. & Beeman, D. (1995) *The Book of GENESIS: Exploring Realistic Neural Models with the GEneral NEural SImulation System.* New York: TELOS/Springer–Verlag.

[3] Hasselmo, M.E., Schnell, E. & Barkai, E. (1995) Dynamics of learning and recall at excitatory recurrent synapses and cholinergic modulation in rat hippocampal region CA3. *Journal of Neuroscience* **15**(7):5249-5262.

# References

Kyunghyun Cho, Bart van Merrienboer, Çaglar Gülçehre, Fethi Bougares, Holger Schwenk, and Yoshua Bengio. Learning phrase representations using RNN encoder-decoder for statistical machine translation. *CoRR*, abs/1406.1078, 2014. URL http://arxiv.org/abs/1406.1078.

Tomas Mikolov, Kai Chen, Greg Corrado, and Jeffrey Dean. Efficient estimation of word representations in vector space. *CoRR*, abs/1301.3781, 2013. URL http://arxiv.org/abs/1301.3781.

M. Schuster and K. K. Paliwal. Bidirectional recurrent neural networks. *IEEE Transactions on Signal Processing*, 45(11):2673–2681, Nov 1997. ISSN 1053-587X. doi: 10.1109/78.650093.

Iulian Vlad Serban, Alessandro Sordoni, Yoshua Bengio, Aaron C. Courville, and Joelle Pineau. Hierarchical neural network generative models for movie dialogues. *CoRR*, abs/1507.04808, 2015. URL http://arxiv.org/abs/1507.04808.

Alessandro Sordoni, Yoshua Bengio, Hossein Vahabi, Christina Lioma, Jakob Grue Simonsen, and Jian-Yun Nie. A hierarchical recurrent encoder-decoder for generative context-aware query suggestion. *CoRR*, abs/1507.02221, 2015. URL http://arxiv.org/abs/1507.02221.

Ilya Sutskever, Oriol Vinyals, and Quoc V. Le. Sequence to sequence learning with neural networks. *CoRR*, abs/1409.3215, 2014. URL http://arxiv.org/abs/1409.3215.