



Eidgenössische Technische Hochschule Zürich
Swiss Federal Institute of Technology Zurich

Report

Genre-aware Seq2seq Model

Stefan Stark
Mohammed Ajil
Vincent Stettler

Contents

1	Motivation	3
2	Model	3
2.1	Bos_genre	3
2.2	Concat_one_hot	3
2.3	Concat_embedding	3
3	Analysis	4
4	Example Sentences	6
5	Existing Literature	8
6	References	8

1 Motivation

The text data we use for sequence learning are sentences from movies. Our idea is to extend a basic seq2seq model such that the information about the genre of movie sentences is taken into account. Our model would then answer questions in the style of a genre. The main motivation behind this approach is to have a simple extension that may give cool results. Unfortunately, the evaluation of this extension is mostly qualitative. However we can compare the overall performance of the extension, qualitatively, with the performance of a basic seq2seq model.

2 Model

In this section we present and discuss three different approaches for including the genre information in a basic seq2seq model. For the evaluation of these models we use an old fashioned seq2seq model as comparison, with no genre information.

2.1 Bos_genre

The first approach is to replace the `<bos>` tag in the decoder sequence with a tag for the genre (e.g., `<comedy>`). These genre tags are included in the vocabulary and hence are encoded as indices, and then embedded. The advantages of this approach are that it is easy to implement, and that it is very cheap in terms of additional model parameters. Apart from the tag embeddings no parameters are added. One potential disadvantage may be the difficulty to have long range dependency. The model might "forget" the information about the genre while decoding long sentences.

2.2 Concat_one_hot

In order to tackle potential problems with long term dependencies, the second model appends a one hot encoding of the genre to the word embedding. The thought behind this approach is that the genre should affect each step of the model. A disadvantage of this model might be the independence among the genres.

2.3 Concat_embedding

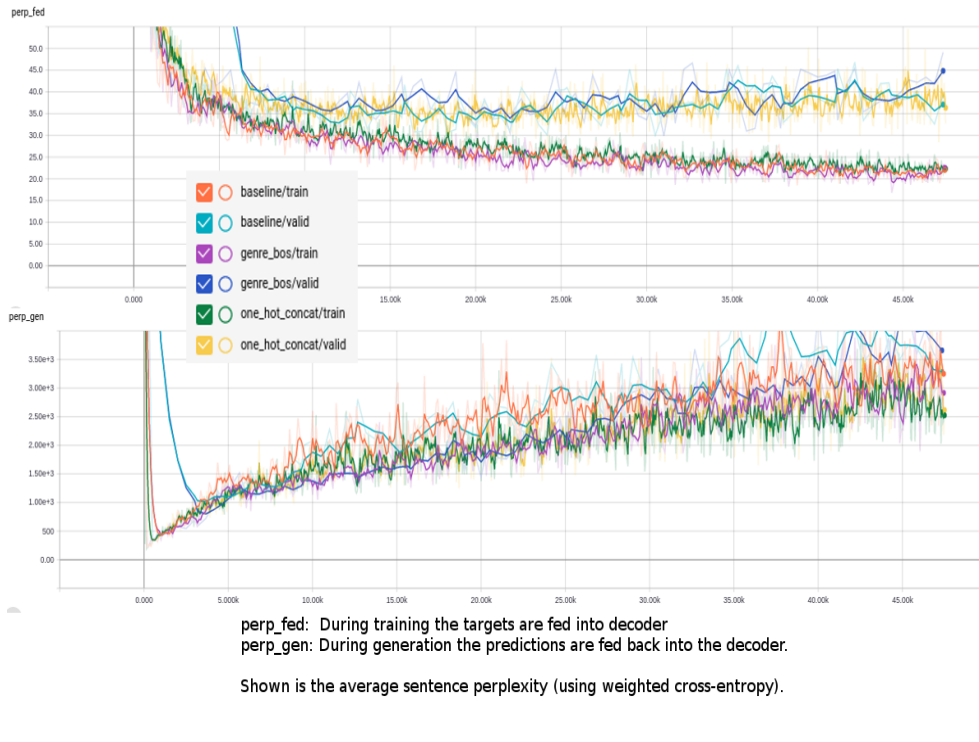
This model embeds the genre before concatenating it to the word embeddings. This should result in the ability to model genre information. Unfortunately,

the genre information is compressed further, while we likely already have a hard time modelling it.

3 Analysis

In this section we discuss the results of our analysis. We implemented a basic seq2seq model, and extended that using the approaches 2.1 and 2.2. Section 6 presents a list of sources that were used to implement the models.

During the training we collected the average sentence perplexity of each batch, for the mode where we feed in the ground truth in the decoder, *and* the mode for sentence generation. Additionally, after every 500 batches, we compute the same quantities for the validation set. In the computation of the perplexities we use a weighted cross-entropy. The following figures visualize these perplexity values during the training.



As one can see in the graphs of the perplexities, our models did unfortunately not perform better than the baseline. This result could be due to the fact that the model already had a hard time learning the language without the genre information. Adding the genre information would then not necessarily lead to an improvement. Further, we do not see any performance

difference between our extensions. We hoped that ensuring the influence of the genre in each step would perform better than the model where we replaced the `<bos>` tag with a genre tag. From this we conclude that either the genre does not have a significant impact on the form of the sentences, or that our implementation of the models was not able to fully use the genre information. It would have been interesting to also test the model described in Section 2.3, where an embedding of the genre would be learned. This would give the model a better representation of the genre tags, since the model will learn the embeddings in a context. These embeddings would only be affected by sentences of the correct genre, therefore containing information about the structure of a dialog in that genre.

Another observation is that the perplexities of the generated words are increasing with the iterations. It makes sense that the perplexities of the generated words are worse than the perplexities of the mode where the targets are fed into the decoder. This is because the decoder, running in feedback mode, could "diverge" away from the correct targets, while the decoder in the other mode is held close to the targets. We do however not know why the perplexities are *increasing*. One possible explanation might be that the models are overfitting.

4 Example Sentences

In this section we present some example sentences of the model that replaces the <bos> tags with genre tags. Each input is answered in multiple genres. (We selected the most interesting answers.)

```
INPUT      > you knew she was pregnant .
<ACTION>   > <person> , i ' m not sure i ' m not . <eos> . <eos> . <eos> . <eos> <eos>
<ADULT>    > i ' m not . <eos> . <eos> . <eos> . <eos> . <eos> . <eos> . <eos> . <eos>
<ANIMATION> > you ' re not . <eos> . <eos> . <eos> . <eos> . <eos> . <eos> . <eos> . <eos>
<COMEDY>   > <person> , i ' m sorry . <eos> . <eos> . <eos> . <eos> . <eos> . <eos> . <eos>
<FAMILY>   > you know , <person> . <eos> . <eos> . <eos> . <eos> . <eos> . <eos> . <eos> . <eos>
<HORROR>   > just to get me a little bit . <eos> . <eos> . <eos> . <eos> . <eos> . <eos> . <eos> <eos>
<MUSIC>    > no . <eos> . <eos> . <eos> . <eos> . <eos> . <eos> . <eos> . <eos> . <eos> <eos>
<MYSTERY>  > to be a lot of <unk> . <eos> . <eos> . <eos> . <eos> . <eos> . <eos> . <eos>
<ROMANCE>  > <person> , i ' m not going to get out of here . <eos> . <eos> . <eos> . <eos>
<SCI-FI>   > . <eos> . <eos> . <eos> . <eos> . <eos> . <eos> . <eos> ? <eos> ? <eos>
<SPORT>    > you were going to be a little . <eos> . <eos> . <eos> . <eos> . <eos> . <eos> <eos>

INPUT      > oh , no . we ' ll never get it out now .
<ACTION>   > you ' re not going to be able to get the <unk> on the <unk> , <person> . <eos> <eos>
<ADULT>    > -- ? <eos> . <eos> . <eos> ? <eos> ? <eos> ? <eos> ? <eos> ? <eos> . <eos> <eos>
<ANIMATION> > <person> , i ' m sorry . <eos> . <eos> . <eos> . <eos> . <eos> . <eos> . <eos>
<FANTASY>  > what ' s the matter ? <eos> . <eos> . <eos> . <eos> . <eos> . <eos> . <eos> <eos>
<HISTORY>  > you want to go ? <eos> . <eos> . <eos> . <eos> . <eos> . <eos> . <eos> . <eos>
<MUSIC>    > you can ' t be . <eos> . <eos> . <eos> . <eos> . <eos> . <eos> . <eos> <eos>
<MYSTERY>  > <person> , i ' m not sure . <eos> . <eos> . <eos> . <eos> . <eos> . <eos> . <eos> <eos>
<ROMANCE>  > you ' re a <unk> ? <eos> . <eos> . <eos> . <eos> . <eos> . <eos> . <eos> <eos>
<SCI-FI>   > <person> , <person> . <eos> . <eos> . <eos> . <eos> . <eos> . <eos> . <eos> . <eos> <eos>
<SPORT>    > you ' re gon na be a little ? <eos> . <eos> . <eos> . <eos> . <eos> . <eos> . <eos>
<THRILLER> > <person> , i ' m sorry . i ' m sorry . <eos> . <eos> . <eos> . <eos> <eos>
<WAR>      > <person> , i ' m not going to be able to see you . <eos> . <eos> . <eos> <eos>
<WESTERN>  > <person> , i ' m sorry . <eos> . <eos> . <eos> . <eos> . <eos> . <eos> . <eos>

INPUT      > eh ? it ' s a four ton truck . it ' s not as though i t ' s a <unk> fuckin a flea is it ?
<ACTION>   > of course it is . <eos> . <eos> . <eos> . <eos> . <eos> . <eos> . <eos> . <eos>
<ADULT>    > for a minute , i ' m not sure . <eos> . <eos> . <eos> . <eos> . <eos> <eos>
<ADVENTURE> > ' s not in the bucks of the night . <eos> . <eos> . <eos> . <eos> . <eos> <eos>
<ANIMATION> > <person> , i ' m sorry . i ' m sorry . <eos> . <eos> . <eos> . <eos> <eos>
<BIOGRAPHY> > ' s a <unk> . <eos> . <eos> . <eos> . <eos> . <eos> . <eos> . <eos> . <eos>
<COMEDY>   > it is . <eos> . <eos> . <eos> . <eos> . <eos> . <eos> . <eos> . <eos>
<CRIME>    > is a <unk> <unk> . <eos> . <eos> . <eos> . <eos> . <eos> . <eos> . <eos> . <eos>
<DOCUMENTARY> > i ' m not sure . <eos> . <eos> . <eos> . <eos> . <eos> . <eos> . <eos> <eos>
<HISTORY>  > not do . <eos> . <eos> . <eos> . <eos> . <eos> . <eos> . <eos> . <eos>
<HORROR>   > <person> , <person> , i ' m sorry , i ' m not a <unk> . <eos> . <eos> <eos>
<MUSIC>    > is <person> ' s a little bit . <eos> . <eos> . <eos> . <eos> . <eos> . <eos> <eos>
<MYSTERY>  > i ' m not sure . <eos> . <eos> . <eos> . <eos> . <eos> . <eos> . <eos> <eos>
<ROMANCE>  > on it . <eos> . <eos> . <eos> . <eos> . <eos> . <eos> . <eos> . <eos>
<SCI-FI>   > i got a lot of money . <eos> . <eos> . <eos> . <eos> . <eos> . <eos> . <eos>
<SHORT>    > . <person> . <eos> . <eos> . <eos> . <eos> . <eos> . <eos> . <eos> . <eos> ? <eos>
<SPORT>    > on the way . <eos> . <eos> . <eos> . <eos> . <eos> . <eos> . <eos> . <eos> <eos>
<WESTERN>  > i ' m sorry . i ' m sorry . <eos> . <eos> . <eos> . <eos> . <eos> <eos>
```

```

INPUT      > do i hear fifty thousand ?
<ACTION>   > <person> ' t worry about it . <eos> . <eos> . <eos> . <eos> . <eos> . <eos> . <eos>
<ANIMATION> > <person> ' t worry about it . <eos> . <eos> . <eos> . <eos> . <eos> . <eos> . <eos>
<COMEDY>   > , <person> . <eos> . <eos> . <eos> . <eos> . <eos> . <eos> . <eos> . <eos> . <eos>
<DOCUMENTARY> > <person> , <person> , i know , but i ' m not sure . <eos> . <eos> . <eos> <eos>
<FAMILY>   > <person> ' t worry about it . <eos> . <eos> . <eos> . <eos> . <eos> . <eos> . <eos>
<FILM-NOIR> > <person> ' t worry about it . <eos> . <eos> . <eos> . <eos> . <eos> . <eos> . <eos>
<HISTORY>  > it was a <unk> . <eos> . <eos> . <eos> . <eos> . <eos> . <eos> . <eos> . <eos>
<MUSIC>    > <person> ' t say ' ' <unk> ' . <eos> . <eos> . <eos> . <eos> . <eos> . <eos>
<MUSICAL>  > and <unk> . <eos> . <eos> . <eos> . <eos> . <eos> . <eos> . <eos> . <eos> . <eos>
<MYSTERY>  > <person> , we ' re not going to be able to the <unk> . <eos> . <eos> . <eos> <eos>
<ROMANCE>  > . <eos> . <eos> . <eos> . <eos> . <eos> ? <eos> ? <eos> ? <eos> ? <eos> ? <eos>
<SCI-FI>   > <person> ' t worry about it . <eos> . <eos> . <eos> . <eos> . <eos> . <eos> . <eos>
<SHORT>    > one . <eos> . <eos> . <eos> . <eos> . <eos> . <eos> . <eos> . <eos> . <eos> <eos>
<SPORT>    > . <eos> . <eos> . <eos> . <eos> . <eos> . <eos> . <eos> ? <eos> ? <eos> ? <eos>
<THRILLER> > <person> ' t worry about it . <eos> . <eos> . <eos> . <eos> . <eos> . <eos> . <eos>

INPUT      > this is a test .
<ADULT>    > and you ' re not going to be a <unk> ? <eos> . <eos> . <eos> . <eos> . <eos>
<ADVENTURE> > . <eos> . <person> . <eos> . <eos> . <eos> . <eos> . <eos> . <eos> . <eos> . <eos>
<ANIMATION> > <person> , <person> . <eos> . <eos> . <eos> . <eos> . <eos> . <eos> . <eos> . <eos> <eos>
<BIOGRAPHY> > . ? <eos> . <person> ? <eos> . <eos> . <eos> . <eos> ? <eos> ? <eos> ? <eos> <eos>
<COMEDY>    > of the <unk> ? <eos> . <eos> . <eos> ? <eos> ? <eos> ? <eos> ? <eos> ? <eos> <eos>
<HISTORY>   > ? <eos> . <eos> . <eos> ? <eos> ? <eos> ? <eos> ? <eos> . <eos> ? <eos> ? <eos>
<HORROR>    > <person> <eos> . <eos> . <eos> . <eos> . <eos> . <eos> . <eos> . <eos> ? <eos> ? <eos>
<MUSIC>     > you ' re not going to be a <unk> ? <eos> . <eos> . <eos> . <eos> . <eos> <eos>
<MUSICAL>   > and a few days . <eos> . <eos> . <eos> . <eos> . <eos> . <eos> . <eos> . <eos>
<MYSTERY>   > <person> ' t you ? <eos> . <eos> . <eos> . <eos> . <eos> . <eos> . <eos> . <eos>
<ROMANCE>   > you ' re not . <eos> . <eos> . <eos> . <eos> . <eos> . <eos> ? <eos> ? <eos>
<SCI-FI>    > <person> . <eos> . <eos> . <eos> . <eos> . <eos> . <eos> . <eos> . <eos> ? <eos> <eos>
<SHORT>     > is a <unk> ? <eos> . <eos> . <eos> . <eos> ? <eos> ? <eos> ? <eos> ? <eos> <eos>
<SPORT>     > you ' re in the bucks of a <unk> ? <eos> . <eos> . <eos> . <eos> . <eos> <eos>

```

5 Existing Literature

David Krueger and Roland Memisevic, in their work on regularizing RNNs (<https://arxiv.org/pdf/1511.08400v7.pdf>) used the adam optimizer with gradient clipping at 200 and a learning rate of .001. We tried to do the same, hoping that we might avoid overfitting.

A Hierarchical Latent Variable Encoder-Decoder Model for Generating Dialogues. Iulian Vlad Serban, Alessandro Sordoni, Ryan Lowe, Laurent Charlin, Joelle Pineau, Aaron Courville, Yoshua Bengio. 2016. <http://arxiv.org/abs/1605.06069>

Building End-To-End Dialogue Systems Using Generative Hierarchical Neural Network Models. Iulian V. Serban, Alessandro Sordoni, Yoshua Bengio, Aaron Courville, Joelle Pineau. 2016. AAAI. <http://arxiv.org/abs/1507.04808>

6 References

<https://github.com/ematvey/tensorflow-seq2seq-tutorials/blob/master/2-seq2seq-advanced.ipynb>

<https://github.com/ematvey/tensorflow-seq2seq-tutorials/blob/master/1-seq2seq.ipynb>

<https://github.com/ematvey/tensorflow-seq2seq-tutorials/blob/master/2-seq2seq-advanced.ipynb>

https://github.com/tensorflow/tensorflow/blob/master/tensorflow/contrib/legacy_seq2seq/python/ops/seq2seq.py