

Controlling Linguistic Style Aspects in Neural Language Generation - Supplementary Materials

Jessica Fidler and Yoav Goldberg

Computer Science Department

Bar-Ilan University

Israel

{jessica.fidler, yoav.goldberg}@gmail.com

Technical details

We implement the conditioned LSTM using the DyNet toolkit.¹

Parameters We use words embedding of size 256 and set the dimension of the model parameters values to 20. Thus, the input vector for the RNN is of size 376. We use the LSTM (Hochreiter and Schmidhuber, 1997) which was proven as effective in many NLP tasks. The LSTM have a dimension of 1024. The output of the LSTM is fed to an MLP which is then used for predicting the next word. The MLP’s hidden layer is also of size 1024.

Training We use mini-batches of size 16, and use the Adam optimizer for 10 iterations over the train-set, where samples are randomly shuffled before each iteration. We choose the model with the highest Perplexity on the development set.

Generation When generating texts, we apply *temperature*. Temperature is commonly used to sharpen the distribution resulting from a softmax layer, and is defined as follows:

$$\text{temperature}(r_t, \tau) = \text{softmax}\left(\frac{r_t}{\tau}\right) \quad (1)$$

Where $\tau \leq 1$ is the temperature parameter. The resulting probabilities are nearly the same for high values of τ , while for low values of τ the probability of the item with the highest expected reward tends to 1. We set τ to 0.6. Then we randomly choose the next token according to the distribution.

Vocabulary To cope with a large vocabulary, we follow Sennrich et al. (2016) and use BPE encoding for words. The BPE (*byte pair encoding*) algorithm (Gage, 1994) represents an open vocabulary through a fixed-size vocabulary by encoding

rare words as sequences of subword units. When decoding, subwords units are unified. We set the vocabulary size to 30,469.

Categorization to Themes

The possible values for the Theme parameter are: *plot*, *acting*, *production*, *effects* and *other*. We collected common words in advance and assigned them to these categories as follows:

Plot story, storytelling, backstory, story-telling, story-line, storyteller, back-story, storytellers, storybook, story-lines, stories, plot, plots, sub-plot, subplots, plotting, plotline, plotted, sub-plot, sub-plots, plotless, plotlines, plothes, plot-holes, plot-wise, plot-line, plot-lines, plot-less, script, scripts, scripted, scripting, scripture, scriptwriter, scriptwriting, scriptwriters, un-scripted, manuscript, well-scripted, scriptures, manuscripts, written, writing, writer, write, writers, well-written, writes, screenwriter, screenwriters, screenwriting, co-writer, underwritten, co-written, writings, Screenwriter, storyline, storylines, tale, tales, fairytale, fairytales, trilogy, trilogies, trilogia, trilogia, trilogie, screenplay, Screenplay, screenplays, characters, character, characterization, characteristics, characterisation, character-driven, characterizations, characteristic, characterized, scenes, scene, scenery

Acting Acting acting, actors, actor, act, acted, actress, acts, actresses, well-acted, actores, acteurs, cast, casting, miscast, casts, casted, well-cast, miscasting, performance, performances, role, roles, played, plays, play, playing, players, player, underplayed, well-played, underplaying

Production director, directors, directorial, directores, directora, directoral, directed, directing, production, productions, post-production, pre-production, co-production

¹<https://github.com/clab/dynet>

Effects effects, playlist, songs, song, singer-songwriter, singer/songwriter, songwriting, songwriters, songwriter, music, special-effects, voice, voices, voiced, voiceover, voice-over, voiceovers, voice-overs, visual, visuals, visually, visuales, visualmente, visualize, visualization, visualized, trailer, trailers, sound, sounds, sounded, sounding, soundtrack, soundtracks, cinematography, shot, shots, well-shot

We categorize a sentence with the theme category with the highest number of the matched words in the category list. If there are no matched words, we categorize the sentence as *other*.

Examples for Generated Sentences

Here are some examples of sentences that were generated by our model with the corresponding parameters values:

Parameter	Value
Theme	Other
Sentiment	Negative
Professional	False
Personal	True
Length	11-20 words
Descriptive	True

- “My biggest problem with the whole movie though is that there is nothing new or original or great in this film.”
- “There are some funny parts but overall I didn’t like the first few funny parts, but overall pretty decent .”
- “My biggest problem with the movie was the fact that is managed to use the same exact same well-written line.”
- “Ultimately, I can honestly say that this movie is full of stupid stupid and stupid stupid stupid stupid stupid.”
- Good but a little bit slow and boring, I was looking forward to seeing this movie with my parents.

Parameter	Value
Theme	Other
Sentiment	Negative
Professional	False
Personal	False
Length	11-20 words
Descriptive	True

- “A little bit of a predictable and boring romantic comedy with a few funny moments but overall pretty entertaining.”
- “With such a great premise ,Escape From Tomorrow is pretty damn terrible, horrible, and no exception.”
- “The first half is lazy and stupid, but there’s a handful of funny moments which are pretty decent.”
- “There’s no denying the fact that this movie is such a horrible movie with a few bad moments.”
- “The last part of the movie just let me down, but the whole thing is pretty good.”
- “It is a little difficult to follow, but this is a rare right choice for the respective aspects of film.”
- “My biggest issue is that the first half is pretty boring, plodding, and too obvious to be honest.”

Parameter	Value
Theme	Plot
Sentiment	Positive
Professional	False
Personal	False
Length	11-20 words
Descriptive	False

- “The movie’s story is based on real life events, and the characters are even more interesting than the original.”
- “The story is great, the most fun and the most great film you’ll see in a long time.”
- “The story is the best part of the movie, it’s a lot of fun to watch.”
- “It ’s a touching story that will keep you on the edge of your seat the whole time ! ! !”
- “The story was not quite as good as the first one but it had a pretty good twist ending.”
- “It’s a story that doesn’t take itself too seriously, but it’s a surprisingly good film.”
- “The movie is a perfect love story that leaves you with a smile on your face for the great effect.”

- “The characters were great and the storyline had me laughing out loud at the beginning of the movie.”
- “The two main characters are a bit of a stretch, but this movie is still very well done.”
- “The film is definitely a unique thing to follow, and some of the characters are just a bit too complicated.”

Parameter	Value
Theme	Other
Sentiment	Positive
Professional	True
Personal	False
Length	11-20 words
Descriptive	False

- “This is a must see for fans of Bergman’s “American Dream”, “The Roots”.”
- “The film’s ultimate pleasure if you want to fall in love with the ending, you won’t be disappointed”
- “One of the most inspirational films I’ve seen in years, hence the most influential, the best.”
- “The monetary system is a bit too intelligent and at times implanted in the middle of the film.”
- “The film doesn’t seem to be anything more than a dozen other films with a sophisticated vigor.”
- “The film’s late 19th century Denmark’s history is a feast for the eyes and the laughter on the screen.”
- “The film’s simple, and a refreshing take on the complex family drama of the regions of human intelligence.”

Parameter	Value
Theme	Acting
Sentiment	Neutral
Professional	False
Personal	False
Length	21-40 words
Descriptive	False

- “It’s a shame to see the actors playing the corporation’s name or the West Memphis Three in the theater, but he still manages to do with a one of a better black comedy.”

- “In the end, the film is a brilliantly acted and terrifically paced adaptation of the dystopian book, but the direction is completely questionable as its subject matter gives us a sense of realism which is a brilliant thing”
- “A good performance from James Franco and Josh Brolin, Peter Dinklage (played by Michael Douglas), his wife and the city of the mayor, who develops into a newly dramatic movie.”

References

- Philip Gage. 1994. A new algorithm for data compression. R & D Publications, Inc., volume 12, pages 23–38.
- Sepp Hochreiter and Jürgen Schmidhuber. 1997. Long short-term memory. MIT Press, volume 9, pages 1735–1780.
- Rico Sennrich, Barry Haddow, and Alexandra Birch. 2016. Neural machine translation of rare words with subword units. In *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*. Association for Computational Linguistics.