Instructions for *ACL Proceedings

Anonymous ACL submission

Abstract

The replicator equations illustrate the evolution of the distribution of types in a population. We utilize replicator dynamics as the backbone to create an architecture, Replicator, which directly learns discrete probability distributions over feature spaces. Replicator could apply whenever needing to learn probability distributions. In Replicator, data features first transform into discrete distributions over a probability space. After that, the discrete distributions evolve according to the replicator equation rules. Finally, the discrete distributions are converted back to the original data space. Replicator achieved the-state-of-art accuracy and performance in both synthetic and realistic datasets.

1 Introduction

2 Approach

2.1 Language Model

Let us define a discrete random variable x on vocabulary X where $X = \{1, \ldots, N\}$. This random variable models the probability of a token occurring in the vocabulary. Therefore, we can write $x_t^i = P_t(x = i)$ represents the probability of a token equal to i at time step t. x_t is the distribution of

Given a sequence of words $(x_1, x_2, \ldots, x_T)$, the goal of a language model is to estimate the joint probability of the sequence:

$$P(x_1, x_2, \ldots, x_T). \tag{1}$$

Since language has a sequential ordering, it is common to factorize the joint probability of the sequence as the product of conditional probabilities:

$$P(x_1, x_2, \ldots, x_T) = \prod_{t=1}^{T} P(x_t | x_1, \ldots, x_{t-1}). \tag{2}$$

Thus, the efficient modeling of conditional probabilities of the next word given all the previous ones, $P(x_t|x_1, \ldots, x_{t-1})$, is critical for the capability of the language model.

Although the neural models computing these conditional probabilities have improved their expressiveness significantly, such as self-attention architectures like the Transformer (Vaswani et al., 2017), they all have one limitation.

We here first using replicator dynamics to directly model the conditional probabilities.

2.2 Replicator Equation

In the field of evolutionary game theory, the replicator equation is defined as follows. Let x_i is the proportion of type i in the population, $(x_1, x_2, \ldots, x_n)$ is the vector of the distribution of types in the population, and $\sum_{i=1}^{n} x_i = 1$. Then the continuous form of replicator equation can be written in the form:

$$\dot{x_i} = x_i \left((Ax)_i - x^T A x \right), \tag{3}$$

where the payoff matrix A holds all the fitness information for the population: the fitness of type i is $(Ax)_i$ and the average fitness of the population is $x^T Ax$.

Similar to the continuous form, the discrete form of the replicator equation can be written as:

$$x_t^i = x_{t-1}^i + x_{t-1}^i \left[(Ax_t)_i - x_t^T A x_t \right]. \tag{4}$$

2.3 Bayesian Inference

2.4 Bayesian Inference is a special case of Replicator Equation

(Harper, 2009)

To the best of our knowledge, however, the Replicator is the first language model relying entirely on the replicator equation to compute the conditional probabilities over the words.

Command	Output	Command	Output
{\"a}	ä	{\c c}	ç
{\^e}	ê	{\u g}	ğ
{\`i}	ì	{\l}	ł
{\.I}	İ	{\~n}	ñ
{\o}	ø	{\H o}	ő
{\'u}	ú	{\v r}	ř
{\aa}	å	{\ss}	ß

Table 1: Example commands for accented characters, to be used in, *e.g.*, BibTeX entries.

3 Model Architecture

Our model does not require position embeddings because the token-mixing Replicator Layer are sensitive to the order of the input tokens.

Spatial information acrosßs tokens (words), which position the words should be.

Our model accepts a word sequence, linearly embeds each word to a feature vector, and obtain a 2-rank tensor with shape (words $\times$ features) as input. Replicator network (architecture) makes use of two types of Replicator layers: word Replicators and feature Replicators. The word Replicators capture co-occurrences patterns among words in the sequence. They operate on each feature independently and take individual columns as inputs. The feature Replicators allow communication between different features.

4 Why Replicator

Dynamic directly in probability space, has no vanishing / exploding gradient problems, thus no normal layer is required.

5 Document Body

5.1 Footnotes

Footnotes are inserted with the `\footnote` command.[1]

5.2 Tables and figures

See Table 1 for an example of a table and its caption.
Do not override the default caption sizes.

5.3 Hyperlinks

Users of older versions of LaTeX may encounter the following error during compilation:

```
\pdfendlink ended up in
different nesting level
than \pdfstartlink.
```

This happens when pdfLaTeX is used and a citation splits across a page boundary. The best way to fix this is to upgrade LaTeX to 2018-12-01 or later.

5.4 Citations

Table 2 shows the syntax supported by the style files. We encourage you to use the natbib styles. You can use the command `\citet` (cite in text) to get "author (year)" citations, like this citation to a paper by Gusfield (1997). You can use the command `\citep` (cite in parentheses) to get "(author, year)" citations (Gusfield, 1997). You can use the command `\citealp` (alternative cite without parentheses) to get "author, year" citations, which is useful for using citations within parentheses (e.g. Gusfield, 1997).

5.5 References

The LaTeX and BibTeX style files provided roughly follow the American Psychological Association format. If your own bib file is named `custom.bib`, then placing the following before any appendices in your LaTeX file will generate the references section for you:

```
\bibliographystyle{acl_natbib}
\bibliography{custom}
```

You can obtain the complete ACL Anthology as a BibTeX file from https://aclweb.org/anthology/anthology.bib.gz. To include both the Anthology and your own .bib file, use the following instead of the above.

```
\bibliographystyle{acl_natbib}
\bibliography{anthology,custom}
```

Please see Section 6 for information on preparing BibTeX files.

5.6 Appendices

Use `\appendix` before any appendix section to switch the section numbering over to letters. See Appendix A for an example.

6 BibTeX Files

Unicode cannot be used in BibTeX entries, and some ways of typing special characters can disrupt BibTeX's alphabetization. The recommended way of typing special characters is shown in Table 1.

[1] This is a footnote.

Output	natbib command	Old ACL-style command
(Gusfield, 1997)	`\citep`	`\cite`
Gusfield, 1997	`\citealp`	no equivalent
Gusfield (1997)	`\citet`	`\newcite`
(1997)	`\citeyearpar`	`\shortcite`

Table 2: Citation commands supported by the style file. The style is based on the natbib package and supports all natbib citation commands. It also supports commands defined in previous ACL style files for compatibility.

Please ensure that BibTeX records contain DOIs or URLs when possible, and for all the ACL materials that you reference. Use the `doi` field for DOIs and the `url` field for URLs. If a BibTeX entry has a URL or DOI field, the paper title in the references section will appear as a hyperlink to the paper, using the hyperref LaTeX package.

Acknowledgements

This document has been adapted by Steven Bethard, Ryan Cotterell and Rui Yan from the instructions for earlier ACL and NAACL proceedings, including those for ACL 2019 by Douwe Kiela and Ivan Vulić, NAACL 2019 by Stephanie Lukin and Alla Roskovskaya, ACL 2018 by Shay Cohen, Kevin Gimpel, and Wei Lu, NAACL 2018 by Margaret Mitchell and Stephanie Lukin, BibTeX suggestions for (NA)ACL 2017/2018 from Jason Eisner, ACL 2017 by Dan Gildea and Min-Yen Kan, NAACL 2017 by Margaret Mitchell, ACL 2012 by Maggie Li and Michael White, ACL 2010 by Jing-Shin Chang and Philipp Koehn, ACL 2008 by Johanna D. Moore, Simone Teufel, James Allan, and Sadaoki Furui, ACL 2005 by Hwee Tou Ng and Kemal Oflazer, ACL 2002 by Eugene Charniak and Dekang Lin, and earlier ACL and EACL formats written by several people, including John Chen, Henry S. Thompson and Donald Walker. Additional elements were taken from the formatting instructions of the *International Joint Conference on Artificial Intelligence* and the *Conference on Computer Vision and Pattern Recognition*.

References

Dan Gusfield. 1997. *Algorithms on Strings, Trees and Sequences*. Cambridge University Press, Cambridge, UK.

Marc Harper. 2009. The replicator equation as an inference dynamic. *arXiv preprint arXiv:0911.1763*.

Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Łukasz Kaiser, and Illia Polosukhin. 2017. Attention is all you need. In *Advances in neural information processing systems*, pages 5998–6008.

A Example Appendix

This is an appendix.