Finding ARG1 of Partitive Nouns in NomBank with DistilBERT

Ziyang Zeng
Dept. of Computer Science
New York University
zz2960@nyu.edu

Abstract

This document is a supplement to the general instructions for *ACL authors. It contains instructions for using the LaTeX style files for ACL conferences. The document itself conforms to its own specifications, and is therefore an example of what your manuscript should look like. These instructions should be used both for papers submitted for review and for final versions of accepted papers.

1 Introduction

NomBank (Meyers et al., 2004c) is a databank project at New York University that adds annotation layer of argument structure for instances in the Penn Treebank II corpus. Apart from the more commonly researched nominalizations of verbs and nominalizations of adjectives, it also covers relational nouns, partitive nouns and several other types of argument-taking nouns.

BERT (Devlin et al., 2018) is a large-size language representation model trained on a large corpus of English text that can achieve state-of-the-art results on a variety of natural language processing tasks. BERT came a decade later than the release of NomBank and there's been little previous work in using it to conduct NomBank-based tasks and related experiments. However, as powerful as BERT is, its large parameters size makes it significantly more computationally expensive to train and use. On the other hand, DistilBERT is a distilled student model of BERT that retains 97% of BERT's language understanding capabilities while being 40% smaller and 60% faster.

This paper presents a token classification approach using pre-trained DistilBERT to find ARG1 of partitive nouns in NomBank. It focuses on partitive nouns (nouns that are used to describe a part or quantity of something) and the partitive task. The task is described as finding one argument (ARG1) of % (the percent sign), or in other words, finding the none group that is being sub-divided or quantified over. For example, for the sentence "Output in the energy sector rose 3.8%.", the ARG1 to be found is "Output" since it is the partitive noun that "3.8%" is referring to. This task can be translated into a binary classification problem on each token in the sentence where the model predicts whether the token is an ARG1 or not.

In addition to the token classification approach above, this paper also experiments with question-answer task, transforming the original ARG1 finding task into the task of make the model answer the question "What is the ARG1 of this sentence?". The advantage of this adaptation is that the model would stably output exactly one ARG1 for each sentence whereas token classification could give each sentence 0 to N ARG1s.

2 Related Work

NomBank (Meyers et al., 2004c), as a databank extending the frames of NOMLEX and PropBank, annotates argument structures of common nouns similar to how PropBank annotates predicating verbs. The development of the NomBank corpus has supported the work of (Jiang and Ng, 2006) and (Liu and Ng, 2007). Each study tested the hypothesis that machine learning methodologies and representations used in verbal SRL (cf. (Pradhan et al., 2005)) can be ported to the task of NomBank SRL. (Liu and Ng, 2007) reports an overall f-measure score of 0.7283 using automatically generated parse trees, demonstrating that modest performance can be achieved using basic approaches developed in the verbal domain. Both studies also investigated the use of features specific to the task of nominal SRL, but observed only marginal performance gains.

3 Preamble

The first line of the file must be

```
\documentclass[11pt]{article}
```

To load the style file in the review version:

```
\usepackage[review]{acl}
```

For the final version, omit the `review` option:

```
\usepackage{acl}
```

To use Times Roman, put the following in the preamble:

```
\usepackage{times}
```

(Alternatives like txfonts or newtx are also acceptable.)

Please see the LaTeX source of this document for comments on other packages that may be useful.

Set the title and author using `\title` and `\author`. Within the author list, format multiple authors using `\and` and `\And` and `\AND`; please see the LaTeX source for examples.

By default, the box containing the title and author names is set to the minimum of 5 cm. If you need more space, include the following in the preamble:

```
\setlength\titlebox{<dim>}
```

where `<dim>` is replaced with a length. Do not set this length smaller than 5 cm.

4   Document Body

4.1   Footnotes

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

4.2   Tables and figures

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

4.3   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.

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.

4.4   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 **?**. You can use the command `\citep` (cite in parentheses) to get "(author, year)" citations (**?**). 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. **?**).

4.5   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 5 for information on preparing BibTeX files.

4.6   Appendices

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

---

[1]This is a footnote.

Output	natbib command	Old ACL-style command
(?)	\citep	\cite
?	\citealp	no equivalent
?	\citet	\newcite
(?)	\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.

5 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.

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

A Example Appendix

This is an appendix.