Vigilante at SemEval-2019 Task 6: How Far Can Models Overcome Data Imbalance

Ioan Budea
Imperial College London
ib2215@ic.ac.uk

Adrian Catană
Imperial College London
ac5915@ic.ac.uk

Abstract

Social media content generation is driven by the constant user growth. With this trend, the amount of hate speech is also increasing, making the necessity of automated methods for offensive language detection more proeminent. This paper presents the empirical results on the OffensEval challenge on identifying and categorizing offensive language in social media. We present the approaches tried, the different architecture of the models and discuss the results obtained and challenges encountered in designing classifiers for this purpose.

1 Introduction

With the emergence of social networking websites, the information posted and shared by users reaches a tremendous audience instantly. However, these websites and technologies have also enabled individuals to disseminate aggressive and harmful content as well, shielded by this layer of indirection. Finding a way to limit the generation and impact of offensive speech has attracted a lot of attention from the research community in recent years (Davidson et al., 2017; Malmasi and Zampieri, 2017). The amount of new data posted every day makes manual monitoring and moderation of content infeasible.

In this paper, we present out findings on the SemEval-2019 Task 6: Identifying and Categorizing Offensive Language in Social Media. The competition required building classifiers for three different subtasks: offensive language Identification (Task A - binary classifier to detect offensive and acceptable language), automatic categorization of offense types (Task B - binary classifier to split the offensive content in Task A into targeted and untargeted) and offense target Identification (Task C - ternary classifier for targeted offensive language). The participants were provided with a dataset of 13200 labelled tweets to be used as training data.

In Natural Language Processing (NLP), classical methods required manual feature engineering and used statistical algorithms to perform the majority of the tasks. The new wave of NLP models borrow ideas from deep learning to perform extraction of hierarchical features from the given data. For our submissions, we decided to use Random Forests as an algorithm representative for the first category, and Convolutional Neural Networks and LSTM for the deep learning approaches.

2 Data Processing

The gievn dataset from the competition starting kit contains 13240 tweets: 4400 of them are marked as offensive and 8840 are not offensive. The offensive ones are further split for task B into 524 untargeted and 3876 targeted. Finally the targeted ones are divided for task C into: ...

We have used the holdout method to split these tweets into a training set 8712 and a validation set comprising the remaining examples.

TODO: After we report the results of our validation experiments, we train the classifiers on the entire dataset before submission.

As a small test set we have parsed an additional file present in the starting kit to construct a dataset of 320 tweets.

We convert the tweets to lowercase, remove the following unnecessary space patterns and non-alphanumeric characters from the tweets, and remove the stopwords using the nltk Python package. After applying data augmentation, stemming or lemmatization (described in the subsections below), a vocabulary is built from the words in the training set. Each word in the sentences is then transformed into a number based on the index in this vocabulary. Then, each sentence is padded to

the maximum length of all sentences by filling the vectors with the element 0 which represents the special ¡pad¿ symbol.

2.1 Data augmentation

From the above ratios, we see that the number of examples from each class are not comparable. While the issue is not so proeminent for the training dataset used for task A, this class imbalance might become problematic when investigating targeted offensive language. Therefore, we have decided to apply the following techniques for data augmentation:

- increasing the number of offensive sample by taking an offensive tweet and appending a random harmless tweet to create a new test instance
- increasing the number of untargeted sample by taking an untargeted tweet and duplicating it as another test instance
- Maybe augment for C?

TODO: explain augmentation after split - risk to have data in validation

2.2 Stemming

We have also attempted word stemming for our tasks. Stemming can be defined as a rather crude heuristic process that removes the affixes of words in the hope of bringing the inflectional forms and derivationally related forms of a word to a common base form. In our work we have used the Porter stemming algorithm implemented in the same nltk Python package.

2.3 Lemmatization

While the goal of lemmatization is similar to the case of stemming, the latter method obtains the stem of a word after applying a set of rules without taking into account the part of speech or the context of the word occurrence. Lemmatization obtains the base form of a word known as a lemma of a word which involves reducing the word forms to its root form after doing a full morphological analysis of a sentence. While this technique may work in some application, it can hurt the performance of others. We have attempted this alternative preprocessing because stemming usually manages to increase recall while harming the precision of the classifiers, so a more informed decision to transform words may provide a better result. For this work, we have used a lemmatizer based on the WordNet lexical database of English words.

2.4 Slang

3 Classifiers

Other work on hate speech in tweets (Badjatiya et al. (2017)) has been conducted using deep learning techniques. Our experiments use two of the methods proposed: Convolutional Neural Networks and Long short-term memory, comparing them against the general framework of Random Forests. WHAT? Moreover, they concluded that embeddings learned from deep neural network models when combined with gradient boosted decision trees led to best accuracy values, which significantly outperforms the existing methods.

In this section we describe the classifiers tested, the framework used for the implementation and the hyperparameters that can be tuned for each approach.

3.1 Random Forest

Random Decision Forests [] are an ensemble learning technique that constructs multiple decision trees using the training data and looks at the mode of the classes to decide the prediction when classifying new data.

The parameters that can be modified are the number of estimators (i.e. the number of trees in the forest), the maximum depth of each tree (having a limit is a good technique for preventing overfitting) and the criterion for each spilt (a choice of either the entropy or gini measure).

Using the RandomForestClassifier from the sklearn Python package, we have built random forests of 30 trees with a maximum depth of 80 nodes for each of the three tasks.

How dafuq we have found the values?

3.2 Convolutional Neural Network

A Convolutional Neural Network (CNN) [] is a feed-forward neural network consisting of the typical input and output layers, but with the hidden layers having additional types: convolution and pooling. The idea of the convolutional layers is to enable the network to automatically identify and extract features from the input space, while pooling layers can be introduced to reduce the dimentionality of the problem space. CNNs are mostly used in applications such as image and

sound recognition, but they can be suited to particular NLP tasks as well. The intuition is that convolutional filters can correspond to several different semantic patterns [Jacovi, 2018] and thus, in our context of offensive language classification, CNNs can extract particular word combinations.

Our CNN is implemented using the PyTorch framework[]. Once the input vectors are built using the index of words in the known vocabulary, these will be used as input to our CNN. The first layer is an embedding layer that transforms each word into a continuous vector space of given dimension (Mikolov). Thus, the word embeddings will be contextualised and learned from the given examples. Next, the output of this embedding layer is fed into a convolutional layer, followed by a ReLU activation function, a max pooling layer and a dropout layer. At the end, a linear layer is applied to bring the result to either a single output value which will be transformed into a probability value for the binary classification task or a vector representing the corresponding outputs of all the classes (i.e. three classes for task C).

The hyperparameters that can be modified are: the number of layers, the dimension of the convolution kernels, the dropout probability, the optimization algorithm and learning rate used.

We have opted for a shallow network architecture because the tweets have a small maximum length, so the number of dimensions doesn't need to be reduced dramatically. By varying the window size from the text taken into account in our convolutional layer we found that ... yields the best results. The CNN was trained using the Adam optimizer [] to accelerate convergence during training with a learning rate of 0.0008 found using grid search.

3.3 Long-Short Term Memory

Recurrent Neural Networks (RNN) have been shown to produce very strong results for language modeling []. They are suited for language tasks because they take into account the whole input sequence, storing the previous results in a hidden state within the nework architecture. Long-short term memory [] is a very popular type of RNNs, proposed to address the problem of vanishing and exploding gradients and make RNN training feasible.

4 Results

Below we present our best runs for the algorithms at each individual task.

4.1 Validation Experiments

100 epochs

RF results will predict only one label CNN with data augmentation, split, stemming: Epoch: 99 — Train Loss: 0.528 — Train Acc: 73.72% — Val. Loss: 0.555 — Val. Acc: 71.25% — — Test Loss: 0.583 — Test Acc: 67.19% 58 Test: Recall: 0.75, Precision: 0.40, F1-measure: 0.52

80 epochs

4.2 Submission Results

4.3 Discussion and Interpretation

Better if we would have used the whole English vocabulary or a bigger subset of it for our word indices since we cannot be ceratin that words in the training set are representative for the test set as well, and this alternative approach would provide a better way of treating new, unseen words

5 Conclusion

6 Credits

This document has been adapted to SemEval-2018 from the instructions for earlier ACL and NAACL proceedings, including those for ACL 2017 by Dan Gildea and Min-Yen Kan, NAACL 2017 by Margaret Mitchell, ACL 2012 by Maggie Li and Michael White, those from ACL 2010 by Jing-Shing Chang and Philipp Koehn, those for ACL 2008 by Johanna D. Moore, Simone Teufel, James Allan, and Sadaoki Furui, those for ACL 2005 by Hwee Tou Ng and Kemal Oflazer, those for ACL 2002 by Eugene Charniak and Dekang Lin, and earlier ACL and EACL formats. Those versions were 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*.

7 Introduction

The following instructions are directed to authors of papers submitted to SemEval-2018 or accepted for publication in its proceedings. All authors are required to adhere to these specifications. Authors

Task A	baseline	stemming	lemmatization
Random Forest	Test Acc: 72.81%	Test Acc: 70.31%	Test Acc: 75.31%
CNN	Test Acc: 82.50% F1-measure: 0.49	Test Acc: 85.94% F1-measure: 0.65	Test Acc: 81.25% F1-measure: 0.41
LSTM	bold	bold	bold

Table 1: Task A experimentation results.

Task B	baseline	stemming	lemmatization
Random Forest	Test Acc: 53.24%	Test Acc: 51.94%	Test Acc: 53.24%
CNN	Test Acc: 68.83% F1-measure: 0.74	Test Acc: 58.44% F1-measure: 0.67	Test Acc: 63.64% F-measure: 0.70
LSTM	bold	bold	bold

Table 2: Task B experimentation results.

are required to provide a Portable Document Format (PDF) version of their papers. **The proceedings are designed for printing on A4 paper.**

SemEval papers should have the exact same format as NAACL-HLT 2018 papers with two exceptions:

1. The review process is single-blind. **Submissions are not anonymous and should use the ACL camera-ready formatting**.

2. **Paper titles should follow the required template.** System description papers submitted by task participants should have a title of the form "[SystemName] at SemEval-2018 Task [TaskNumber]: [Insert Paper Title Here]". Task description papers submitted by task organizers should have a title of the form "SemEval-2018 Task [TaskNumber]: [Task Name]".

8 General Instructions

Manuscripts must be in two-column format. Exceptions to the two-column format include the title, authors' names and complete addresses, which must be centered at the top of the first page, and any full-width figures or tables (see the guidelines in Subsection 8.6). **Type single-spaced.** Start all pages directly under the top margin. See the guidelines later regarding formatting the first page. The manuscript should be printed single-sided and its length should not exceed the maximum page limit described in Section 10. Pages are numbered for initial submission. However, **do not number the pages in the camera-ready version**.

By uncommenting `\aclfinalcopy` at the top of this document, it will compile to produce an example of the camera-ready formatting; by leaving it commented out, the document will be anonymized for initial submission. When you first create your submission on softconf, please fill in your submitted paper ID where `***` appears in the `\def\aclpaperid{***}` definition at the top.

The review process is double-blind, so do not include any author information (names, addresses) when submitting a paper for review. However, you should maintain space for names and addresses so that they will fit in the final (accepted) version. The NAACL-HLT 2018 LaTeX style will create a titlebox space of 2.5in for you when `\aclfinalcopy` is commented out.

The author list for submissions should include all (and only) individuals who made substantial contributions to the work presented. Each author listed on a submission to NAACL-HLT 2018 will be notified of submissions, revisions and the final decision. No authors may be added to or removed from submissions to NAACL-HLT 2018 after the submission deadline.

8.1 The Ruler

The NAACL-HLT 2018 style defines a printed ruler which should be presented in the version submitted for review. The ruler is provided in order that reviewers may comment on particular lines in the paper without circumlocution. If you are preparing a document without the provided style files, please arrange for an equivalent ruler to appear on the final output pages. The presence or absence of the ruler should not change the appearance of any other content on the page. The camera ready copy should not contain a ruler. (LaTeX users may uncomment the `\aclfinalcopy` command in

the document preamble.)

Reviewers: note that the ruler measurements do not align well with lines in the paper – this turns out to be very difficult to do well when the paper contains many figures and equations, and, when done, looks ugly. In most cases one would expect that the approximate location will be adequate, although you can also use fractional references (*e.g.*, the first paragraph on this page ends at mark 108.5).

8.2 Electronically-available resources

NAACL-HLT provides this description in LaTeX2e (`naaclhlt2018.tex`) and PDF format (`naaclhlt2018.pdf`), along with the LaTeX2e style file used to format it (`naaclhlt2018.sty`) and an ACL bibliography style (`acl_natbib.bst`) and example bibliography (`naaclhlt2018.bib`). These files are all available at `http://naacl2018.org/downloads/naaclhlt2018-latex.zip`. We strongly recommend the use of these style files, which have been appropriately tailored for the NAACL-HLT 2018 proceedings.

8.3 Format of Electronic Manuscript

For the production of the electronic manuscript you must use Adobe's Portable Document Format (PDF). PDF files are usually produced from LaTeX using the *pdflatex* command. If your version of LaTeX produces Postscript files, you can convert these into PDF using *ps2pdf* or *dvipdf*. On Windows, you can also use Adobe Distiller to generate PDF.

Please make sure that your PDF file includes all the necessary fonts (especially tree diagrams, symbols, and fonts with Asian characters). When you print or create the PDF file, there is usually an option in your printer setup to include none, all or just non-standard fonts. Please make sure that you select the option of including ALL the fonts. **Before sending it, test your PDF by printing it from a computer different from the one where it was created.** Moreover, some word processors may generate very large PDF files, where each page is rendered as an image. Such images may reproduce poorly. In this case, try alternative ways to obtain the PDF. One way on some systems is to install a driver for a postscript printer, send your document to the printer specifying "Output to a file", then convert the file to PDF.

It is of utmost importance to specify the **A4 format** (21 cm x 29.7 cm) when formatting the paper. When working with `dvips`, for instance, one should specify `-t a4`. Or using the command `\special{papersize=210mm,297mm}` in the latex preamble (directly below the `\usepackage` commands). Then using `dvipdf` and/or `pdflatex` which would make it easier for some.

Print-outs of the PDF file on A4 paper should be identical to the hardcopy version. If you cannot meet the above requirements about the production of your electronic submission, please contact the publication chairs as soon as possible.

8.4 Layout

Format manuscripts two columns to a page, in the manner these instructions are formatted. The exact dimensions for a page on A4 paper are:

- Left and right margins: 2.5 cm
- Top margin: 2.5 cm
- Bottom margin: 2.5 cm
- Column width: 7.7 cm
- Column height: 24.7 cm
- Gap between columns: 0.6 cm

Papers should not be submitted on any other paper size. If you cannot meet the above requirements about the production of your electronic submission, please contact the publication chairs above as soon as possible.

8.5 Fonts

For reasons of uniformity, Adobe's **Times Roman** font should be used. In LaTeX2e this is accomplished by putting

```
\usepackage{times}
\usepackage{latexsym}
```

in the preamble. If Times Roman is unavailable, use **Computer Modern Roman** (LaTeX2e's default). Note that the latter is about 10% less dense than Adobe's Times Roman font.

Type of Text	Font Size	Style
paper title	15 pt	bold
author names	12 pt	bold
author affiliation	12 pt	
the word "Abstract"	12 pt	bold
section titles	12 pt	bold
document text	11 pt	
captions	10 pt	
abstract text	10 pt	
bibliography	10 pt	
footnotes	9 pt	

Table 3: Font guide.

Command	Output
{\"a}	ä
{\^e}	ê
{\'i}	ì
{\.I}	İ
{\o}	ø
{\'u}	ú
{\aa}	å

Command	Output
{\c c}	ç
{\u g}	ğ
{\l}	ł
{\~n}	ñ
{\H o}	ő
{\v r}	ř
{\ss}	ß

Table 4: Example commands for accented characters, to be used in, *e.g.*, BIBTEX names.

8.6 The First Page

Center the title, author's name(s) and affiliation(s) across both columns. Do not use footnotes for affiliations. Do not include the paper ID number assigned during the submission process. Use the two-column format only when you begin the abstract.

Title: Place the title centered at the top of the first page, in a 15-point bold font. (For a complete guide to font sizes and styles, see Table 3) Long titles should be typed on two lines without a blank line intervening. Approximately, put the title at 2.5 cm from the top of the page, followed by a blank line, then the author's names(s), and the affiliation on the following line. Do not use only initials for given names (middle initials are allowed). Do not format surnames in all capitals (*e.g.*, use "Mitchell" not "MITCHELL"). Do not format title and section headings in all capitals as well except for proper names (such as "BLEU") that are conventionally in all capitals. The affiliation should contain the author's complete address, and if possible, an electronic mail address. Start the body of the first page 7.5 cm from the top of the page.

The title, author names and addresses should be completely identical to those entered to the electronical paper submission website in order to maintain the consistency of author information among all publications of the conference. If they are different, the publication chairs may resolve the difference without consulting with you; so it is in your own interest to double-check that the information is consistent.

Abstract: Type the abstract at the beginning of the first column. The width of the abstract text should be smaller than the width of the columns for the text in the body of the paper by about 0.6 cm on each side. Center the word **Abstract** in a 12 point bold font above the body of the abstract. The abstract should be a concise summary of the general thesis and conclusions of the paper. It should be no longer than 200 words. The abstract text should be in 10 point font.

Text: Begin typing the main body of the text immediately after the abstract, observing the two-column format as shown in

the present document. Do not include page numbers.

Indent: Indent when starting a new paragraph, about 0.4 cm. Use 11 points for text and subsection headings, 12 points for section headings and 15 points for the title.

8.7 Sections

Headings: Type and label section and subsection headings in the style shown on the present document. Use numbered sections (Arabic numerals) in order to facilitate cross references. Number subsections with the section number and the subsection number separated by a dot, in Arabic numerals. Do not number subsubsections.

Citations: Citations within the text appear in parentheses as (**?**) or, if the author's name appears in the text itself, as Gusfield (**?**). Using the provided LATEX style, the former is accomplished using `\cite` and the latter with `\shortcite` or `\newcite`. Collapse multiple citations as in (**??**); this is accomplished with the provided style using commas within the `\cite` command, *e.g.*, `\cite{Gusfield:97,Aho:72}`. Append lowercase letters to the year in cases of ambiguities. Treat double authors as in (**?**), but write as in (**?**) when more than two authors are involved. Collapse multiple citations as in (**??**). Also refrain from using full citations as sentence constituents.

We suggest that instead of

"(**?**) showed that ..."

output	natbib	previous ACL style files
(**?**)	`\citep`	`\cite`
?	`\citet`	`\newcite`
(**?**)	`\citeyearpar`	`\shortcite`

Table 5: Citation commands supported by the style file. The citation 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.

you use

"Gusfield (**?**) showed that ..."

If you are using the provided LaTeX and BibTeX style files, you can use the command `\citet` (cite in text) to get "author (year)" citations.

If the BibTeX file contains DOI fields, the paper title in the references section will appear as a hyperlink to the DOI, using the hyperref LaTeX package. To disable the hyperref package, load the style file with the `nohyperref` option:
`\usepackage[nohyperref]{naaclhlt2018}`

Digital Object Identifiers: As part of our work to make ACL materials more widely used and cited outside of our discipline, ACL has registered as a CrossRef member, as a registrant of Digital Object Identifiers (DOIs), the standard for registering permanent URNs for referencing scholarly materials. As of 2017, we are requiring all camera-ready references to contain the appropriate DOIs (or as a second resort, the hyperlinked ACL Anthology Identifier) to all cited works. Thus, please ensure that you use BibTeX records that contain DOI or URLs for any of the ACL materials that you reference. Appropriate records should be found for most materials in the current ACL Anthology at `http://aclanthology.info/`.

As examples, we cite (**?**) to show you how papers with a DOI will appear in the bibliography. We cite (**?**) to show how papers without a DOI but with an ACL Anthology Identifier will appear in the bibliography.

As reviewing will be double-blind, the submitted version of the papers should not include the authors' names and affiliations. Furthermore, self-references that reveal the author's identity, *e.g.*,

"We previously showed (**?**) ..."

should be avoided. Instead, use citations such as

"**?** (**?**) previously showed ... "

Any preliminary non-archival versions of submitted papers should be listed in the submission form but not in the review version of the paper. NAACL-HLT 2018 reviewers are generally aware that authors may present preliminary versions of their work in other venues, but will not be provided the list of previous presentations from the submission form.

Please do not use anonymous citations and do not include when submitting your papers. Papers that do not conform to these requirements may be rejected without review.

References: Gather the full set of references together under the heading **References**; place the section before any Appendices, unless they contain references. Arrange the references alphabetically by first author, rather than by order of occurrence in the text. By using a .bib file, as in this template, this will be automatically handled for you. See the `\bibliography` commands near the end for more.

Provide as complete a citation as possible, using a consistent format, such as the one for *Computational Linguistics* or the one in the *Publication Manual of the American Psychological Association* (**?**). Use of full names for authors rather than initials is preferred. A list of abbreviations for common computer science journals can be found in the ACM *Computing Reviews* (**?**).

The LaTeX and BibTeX style files provided roughly fit the American Psychological Association format, allowing regular citations, short citations and multiple citations as described above.

- Example citing an arxiv paper: (**?**).
- Example article in journal citation: (**?**).
- Example article in proceedings, with location: (**?**).
- Example article in proceedings, without location: (**?**).

See corresponding .bib file for further details.

Submissions should accurately reference prior and related work, including code and data. If a

piece of prior work appeared in multiple venues, the version that appeared in a refereed, archival venue should be referenced. If multiple versions of a piece of prior work exist, the one used by the authors should be referenced. Authors should not rely on automated citation indices to provide accurate references for prior and related work.

Appendices: Appendices, if any, directly follow the text and the references (but see above). Letter them in sequence and provide an informative title: **Appendix A. Title of Appendix**.

8.8 Footnotes

Footnotes: Put footnotes at the bottom of the page and use 9 point font. They may be numbered or referred to by asterisks or other symbols.[1] Footnotes should be separated from the text by a line.[2]

8.9 Graphics

Illustrations: Place figures, tables, and photographs in the paper near where they are first discussed, rather than at the end, if possible. Wide illustrations may run across both columns. Color illustrations are discouraged, unless you have verified that they will be understandable when printed in black ink.

Captions: Provide a caption for every illustration; number each one sequentially in the form: "Figure 1. Caption of the Figure." "Table 1. Caption of the Table." Type the captions of the figures and tables below the body, using 11 point text.

8.10 Accessibility

In an effort to accommodate people who are colorblind (as well as those printing to paper), grayscale readability for all accepted papers will be encouraged. Color is not forbidden, but authors should ensure that tables and figures do not rely solely on color to convey critical distinctions. A simple criterion: All curves and points in your figures should be clearly distinguishable without color.

9 Translation of non-English Terms

It is also advised to supplement non-English characters and terms with appropriate transliterations and/or translations since not all readers understand all such characters and terms. Inline transliteration or translation can be represented in the order of: original-form transliteration "translation".

[1]This is how a footnote should appear.

[2]Note the line separating the footnotes from the text.

10 Length of Submission

The NAACL-HLT 2018 main conference accepts submissions of long papers and short papers. Long papers may consist of up to eight (8) pages of content plus unlimited pages for references. Upon acceptance, final versions of long papers will be given one additional page – up to nine (9) pages of content plus unlimited pages for references – so that reviewers' comments can be taken into account. Short papers may consist of up to four (4) pages of content, plus unlimited pages for references. Upon acceptance, short papers will be given five (5) pages in the proceedings and unlimited pages for references.

For both long and short papers, all illustrations and tables that are part of the main text must be accommodated within these page limits, observing the formatting instructions given in the present document. Supplementary material in the form of appendices does not count towards the page limit; see appendix A for further information.

However, note that supplementary material should be supplementary (rather than central) to the paper, and that reviewers may ignore supplementary material when reviewing the paper (see Appendix A). Papers that do not conform to the specified length and formatting requirements are subject to be rejected without review.

Workshop chairs may have different rules for allowed length and whether supplemental material is welcome. As always, the respective call for papers is the authoritative source.

Acknowledgments

The acknowledgments should go immediately before the references. Do not number the acknowledgments section. Do not include this section when submitting your paper for review.

Preparing References:

Include your own bib file like this:
`\bibliographystyle{acl_natbib}`
`\bibliography{naaclhlt2018}`

Where `naaclhlt2018` corresponds to a naaclhlt2018.bib file.

A Supplemental Material

Submissions may include resources (software and/or data) used in in the work and described

in the paper. Papers that are submitted with accompanying software and/or data may receive additional credit toward the overall evaluation score, and the potential impact of the software and data will be taken into account when making the acceptance/rejection decisions. Any accompanying software and/or data should include licenses and documentation of research review as appropriate.

NAACL-HLT 2018 also encourages the submission of supplementary material to report preprocessing decisions, model parameters, and other details necessary for the replication of the experiments reported in the paper. Seemingly small preprocessing decisions can sometimes make a large difference in performance, so it is crucial to record such decisions to precisely characterize state-of-the-art methods.

Nonetheless, supplementary material should be supplementary (rather than central) to the paper. **Submissions that misuse the supplementary material may be rejected without review.** Essentially, supplementary material may include explanations or details of proofs or derivations that do not fit into the paper, lists of features or feature templates, sample inputs and outputs for a system, pseudo-code or source code, and data. (Source code and data should be separate uploads, rather than part of the paper).

The paper should not rely on the supplementary material: while the paper may refer to and cite the supplementary material and the supplementary material will be available to the reviewers, they will not be asked to review the supplementary material.

Appendices (*i.e.* supplementary material in the form of proofs, tables, or pseudo-code) should be **uploaded as supplementary material** when submitting the paper for review. Upon acceptance, the appendices come after the references, as shown here. Use `\appendix` before any appendix section to switch the section numbering over to letters.

B Multiple Appendices

. . . can be gotten by using more than one section. We hope you won't need that.