Shift-Invariant Dictionary Learning using TCN-WTA Autoencoders for Discovering Musical Relations

Anonymous ACL submission

Abstract

Music hierarchical temporal structure is full of shift invariant patters. The standard methods to encode a generic sequence is usually achieved by recurrent architectures or more recently with transformer that adopts the mechanism of attention. However, RNNs and transformers models do not take advantage of this prior information, or attempt to find repetitive building blocks. Temporal Convolutional Nets can be used to extract shift invariant features of a specific length defined by the kernel size. Using a fully convolutional temporal autoencoder we can find a shift invariant dictionary that can recreate multivariate musical signals. This architecture can strided with no overlap, and be combined to K-WTA activation function to obtain a sparse dictionary promote a sparse representation. In addition to gaining insight into this shift invariant patters, some results indicate that CNN architectures can outperform recurrent networks on specific task and provide server other advantages across a diverse range of tasks and datasets, while demonstrating longer effective memory. We show a few applications of this sparse representation on task to find key signatures, time signatures, artist detection, and music generation. To assist related work, we have made code available.

1 Introduction

What are the benefits of having sparse models? First, as we will show, they can be used to encode prior knowledge in the sparsity patterns. Second, they are lightweight— requiring less memory to store and allowing faster inference and easier interpretability. Nowadays, we often start with models with hundreds of millions to billions of parameters. Sparsity provides a way to completely discard some of these parameters in an informed and principled manner, resulting in smaller model size. For example, mobile applications (e.g., Google Now, Siri, etc.) stand to benefit from smaller models since mo- bile phones typically have less storage and computing power than standard computers. Sparse models come with their own challenges. New varieties of sparse models often require a specialized optimization method, as we will see throughout this thesis. Last, some of the state-of-the-art methods for benchmarks tasks in various application areas such as computational biology (Kim and Xing, 2008) and computer vision are sparse models (Ranzato et al., 2006; Lin and Kung, 2014), empirically demonstrating that they can also lead to statistical improvements if the prior knowledge is correct (Stojnic et al., 2009).

2 Background

The following instructions are directed to authors of papers submitted to NLP4MusA or accepted for publication in its proceedings. All authors are required to adhere to these specifications. Authors are required to provide a Portable Document Format (PDF) version of their papers. **The proceedings are designed for printing on A4 paper.**

3 Shift-Invariant Dictionary Learning

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 3.6). **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 4.1. 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 NLP4MusA LaTeX style will create a titlebox space of 2.5in for you when `\aclfinalcopy` is commented out.

3.1 Temporal Convolutional Networks

The NLP4MusA 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).

3.2 K-WTA Autoencoders

NLP4MusA provides this description in LaTeX2e (`nlp4MusA.tex`) and PDF format (`nlp4MusA.pdf`), along with the LaTeX2e style file used to format it (`nlp4MusA.sty`) and an ACL bibliography style (`nlp4MusA_natbib.bst`) and example bibliography (`nlp4MusA.bib`). These files are all available at `http://blabla.../nlp4MusA-latex.zip`. We strongly recommend the use of these style files, which have been appropriately tailored for the NLP4MusA 2020 proceedings.

3.3 Sparse Representations

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.

3.4 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

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
subsection titles	11 pt	bold
document text	11 pt	
captions	10 pt	
abstract text	10 pt	
bibliography	10 pt	
footnotes	9 pt	

Table 1: Font guide.

- 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.

3.5 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.

3.6 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 1) 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 "CNN") 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.

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 2: Example commands for accented characters, to be used in, *e.g.*, BibTeX names.

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.

3.7 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 (Gusfield, 1997) or, if the author's name appears in the text itself, as Gusfield (1997). Using the provided LaTeX style, the former is accomplished using `\cite` and the latter with `\shortcite` or `\newcite`. Collapse multiple citations as in (Gusfield, 1997; Aho and Ullman, 1972); 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 (Aho and Ullman, 1972), but write as in (Chandra et al., 1981) when more than two authors are involved. Collapse multiple citations as in (Gusfield, 1997; Aho and Ullman, 1972). Also refrain from using full citations as sentence constituents.

We suggest that instead of

> "(Gusfield, 1997) showed that ..."

you use

> "Gusfield (1997) 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.

You can use the command `\citealp` (alternative cite without parentheses) to get "author year" citations (which is useful for using citations within parentheses, as in Gusfield, 1997).

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]{nlp4MusA}`

Compilation Issues: Some of you might 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. To fix this, the style file contains a patch consisting of the following two lines: (1) `\RequirePackage{etoolbox}` (line 454 in `nlp4MusA.sty`), and (2) A long line below (line 455 in `nlp4MusA.sty`).

If you still encounter compilation issues even with the patch enabled, disable the patch by commenting the two lines, and then disable the `hyperref` package (see above), recompile and see the problematic citation. Next rewrite that sentence containing the citation. (See, *e.g.*, `http://tug.org/errors.html`)

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. 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* (American Psychological Association, 1983). 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* (for Computing Machinery, 1983).

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: (Rasooli and Tetreault, 2015).
- Example article in journal citation: (Ando and Zhang, 2005).
- Example article in proceedings, with location: (Borschinger and Johnson, 2011).
- Example article in proceedings, without location: (Andrew and Gao, 2007).

See corresponding .bib file for further details.

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

Dataset	Size	Intrumentation	Encoding
MAESTRO (Gusfield, 1997)	1256	piano	piano
Groove Gusfield (1997)	1200	drum	piano

Table 3: Datasets used to experiment with fully convolutional temporal autoencoder model. All datasets used are MIDI format

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**.

3.8 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]

3.9 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 10 point text. Captions should be placed below illustrations. Captions that are one line are centered (see Table 1). Captions longer than one line are left-aligned (see Table 2). Do not overwrite the default caption sizes. The nlp4MusA.sty file is compatible with the caption and subcaption packages; do not add optional arguments.

3.10 Accessibility

In an effort to accommodate people who are color-blind (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.

[1]This is how a footnote should appear.

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

4 Experiemnt

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".

4.1 Datasets

The NLP4MusA main conference accepts submissions of extended abstracts and short papers. Extended abstracts may consist of up to two (2) pages of content plus unlimited pages for references. Upon acceptance, final versions of extended abstracts will be given no additional page. Short papers may consist of up to four (4) pages of content, plus unlimited pages for references. Upon acceptance, short papers will be given no additional page. For both extended abstracts 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. Papers that do not conform to the specified length and formatting requirements are subject to be rejected without review.

NLP4MisA does encourage the submission of additional material that is relevant to the reviewers but not an integral part of the paper. There are two such types of material: appendices, which can be read, and non-readable supplementary materials, often data or code. Do not include this additional material in the same document as your main paper. Additional material must be submitted as one or more separate files, and must adhere to the same anonymity guidelines as the main paper. The paper must be self-contained: it is optional for re-

viewers to look at the supplementary material. Papers should not refer, for further detail, to documents, code or data resources that are not available to the reviewers. Refer to Appendix A and Appendix B for further information.

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{nlp4MusA_natbib}`
`\bibliography{nlp4MusA}`
where `nlp4MusA` corresponds to a nlp4MusA.bib file.

References

Alfred V. Aho and Jeffrey D. Ullman. 1972. *The Theory of Parsing, Translation and Compiling*, volume 1. Prentice-Hall, Englewood Cliffs, NJ.

American Psychological Association. 1983. *Publications Manual*. American Psychological Association, Washington, DC.

Rie Kubota Ando and Tong Zhang. 2005. A framework for learning predictive structures from multiple tasks and unlabeled data. *Journal of Machine Learning Research*, 6:1817–1853.

Galen Andrew and Jianfeng Gao. 2007. Scalable training of L1-regularized log-linear models. In *Proceedings of the 24th International Conference on Machine Learning*, pages 33–40.

Benjamin Borschinger and Mark Johnson. 2011. A particle filter algorithm for Bayesian wordsegmentation. In *Proceedings of the Australasian Language Technology Association Workshop 2011*, pages 10–18, Canberra, Australia.

Ashok K. Chandra, Dexter C. Kozen, and Larry J. Stockmeyer. 1981. Alternation. *Journal of the Association for Computing Machinery*, 28(1):114–133.

Association for Computing Machinery. 1983. *Computing Reviews*, 24(11):503–512.

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

Mohammad Sadegh Rasooli and Joel R. Tetreault. 2015. Yara parser: A fast and accurate dependency parser. *Computing Research Repository*, arXiv:1503.06733. Version 2.

A Appendices

Appendices are material that can be read, and include lemmas, formulas, proofs, and tables that are not critical to the reading and understanding of the paper. Appendices 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 Supplemental Material

Submissions may include non-readable supplementary material used in the work and described in the paper. Any accompanying software and/or data should include licenses and documentation of research review as appropriate. Supplementary material may 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.** 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.