|  |
| --- |
| **Music Genre Recognition Using Temporal Features** |

**Jivjot SinghMangeshBhangare**

Department of Computing Science Department of Computer Science

Simon Fraser University Simon Fraser University

[*jivjots@sfu.ca*](mailto:jivjots@sfu.ca)[*mbhangar@sfu.ca*](mailto:mbhangar@sfu.ca)

**Anuj Issar Bikramdeep Singh**

Department of Computing Science Department of Computer Science

Simon Fraser University Simon Fraser University

[*aissar@sfu.ca*](mailto:aissar@sfu.ca)[*bsa61@sfu.ca*](mailto:bsa61@sfu.ca)

**Abstract**

This project presents an approach for the music genre classification problem. The proposed approach uses temporal feature vector and weighted voting to improve the prediction accuracy. Classical machine learning algorithms such as Naïve-Bayes, k Nearest-Neighbors, and Support Vector Machines are employed and weighted voting procedures were employed in order to enhance final prediction results. Experiments were carried out on a dataset obtained from Music Analysis, Retrieval and Synthesis for Audio Signals (MARSYAS) which is an open source software framework and data collection of audio files. The dataset contains 1,000 audio files categorized in 10 musical genres. Experimental results show that the proposed ensemble approach produces better results than the ones obtained from individual classifiers.

**1 Introduction**

Duke Ellington who once very wisely said: There are simply two kinds of music, good music and the other kind. The only yardstick by which the result should be judged is simply that of how it sounds. If it sounds good, it's successful; if it doesn't it has failed. Exploring different genres of music is about wanting to know a little bit more about some of the things we humans can do, the feelings we didn't want to leave unsaid; the messages we’ve wanted to get across. Today with the amount of music we have, automatic procedures capable of dealing with large amounts of music in digital formats are imperative, and Music Information Retrieval (MIR) has become an important research area. An important task in MIR is Music Genre Classification problem, music genres are categorical labels created by experts in order to identify the style of the music. The music genre is a descriptor that is largely used to organize collections of digital music. It is not only a crucial metadata in large music databases and electronic music distribution (EMD) [1].

The music can be considered as a high-dimensional digital time-variant signal and considering the amount of music data we have today, it is a good opportunity to automate music genre classification using temporal feature vectors. Marsyas (Music Analysis, Retrieval and Synthesis for Audio Signals) is an open source software framework for audio processing with specific emphasis on Music Information Retrieval applications. For feature extraction we are using Marsyas’s bextract library. [2]\* ‘Bextract’ is executable provided by Marsyas which can be used for complete feature extraction and classification experiments with multiple files.

The approach involves classical machine algorithms such as Naïve-Bayes, k Nearest-Neighbors, and Support Vector Machines and using weighted voting procedures to improve final prediction results. The Naïve Bayes is implemented with 10-fold cross validation using ‘e1071’ package in R. For Support Vector Machines, we have implemented ‘libSVM’ with 10-fold cross validation. The K-Nearest Neighbor is implemented with ‘kkNN’ package of R.

**1.1 Style**

Papers to be submitted to NIPS 2015 must be prepared according to the instructions presented here. Papers may be only up to 8 pages long, including figures and references. Since 2009, an additional ninth page *containing only cited references* is allowed. Papers that exceed nine pages will not be reviewed, or in any other way considered for presentation at the conference.

Please note that this year we have introduced automatic line number generation into the style file (for LaTeX 2e and MS Word versions). This is to help reviewers refer to specific lines of the paper when they make their comments. Please do NOT refer to these line numbers in your paper as they will be removed from the style file for the final version of accepted papers.

The margins in 2015 are the same as since 2007, which allow for ~ 15% more words in the paper compared to earlier years. We are also again using double-blind reviewing. Both of these require the use of new style files.

Authors are required to use the NIPS LaTeX or RTF (MS Word) style files obtainable at the NIPS website as indicated below. Please make sure you use the current files and not previous versions. Tweaking the style files may be grounds for rejection.

**1.2 Retrieval of style files**

The style ﬁles for NIPS and other conference information are available on the World Wide Web at

http://www.nips.cc/

The file **nips2015.pdf** contains these instructions and illustrates the various formatting requirements that your NIPS paper must satisfy. LaTeX users can choose between two style files: **nips11submit\_09.sty** (to be used with LaTeX version 2.09) and **nips11submit\_e.sty** (to be used with LaTeX2e). The file **nips2015.tex** may be used as a “shell” for writing your paper. All you have to do is replace the author, title, abstract and text of the paper with your own. The file **nips2015.rtf** is provided as a shell for MS Word users.

The formatting instructions contained in these style files are summarized in sections 2, 3, and 4, below.

**2 General formatting instructions**

The text must be confined within a rectangle 5.5 inches (33 picas) wide and 9 inches (54 picas) long. The left margin is 1.5 inches (9 picas). Use 10 point type with a vertical spacing of 11 points. Times New Roman is the preferred typeface throughout. Paragraphs are separated by ½ line space, with no indentation.

Paper title is 17 point, initial caps/lower case, bold, centered between 2 horizontal rules. Top rule is 4 points thick and bottom rule is 1 point thick. Allow ¼ inch space above and below title to rules. All pages should start 1 inch (6 picas) from the top of the page.

The version of the paper submitted for review should have "Anonymous Author(s)" as the author of the paper. For the final version, authors’ names are set in boldface, and each name is centered above the corresponding address. The lead author’s name is to be listed first (left-most), and the co-authors’ names (if different address) are set to follow. If only one co-author, list both author and co-author side by side.

Please pay special attention to the instructions in section 4 regarding figures, tables, acknowledgements, and references.

**3 Techniques**

**3.1 K-nearest neighbors**

The first machine learning technique used is the simple K nearest neighbor. We used R library to implement the classifier. The results by the standard R library were not satisfactory, and we tried the new KKNN, a library for weighted k-nearest neighbors classification.

The KKNN uses the Minkowski distance to calculate the nearest neighbors of the given data point. The library provides various options for kernels implementation like "triweight", "cos", "inv", "gaussian", our results were best for the Gaussian kernel. We created a script to determine the best kernel and the optimum number of neighbors for the implementation. The bet results were generated by using a Gaussian kernel with five neighbors.

**3.2 SVM**

Support Vector machines (SVM) are considered one of the best machine learning algorithm for multiclass classifications along with the neural networks. Our dataset being small, SVM classifier was expected to perform best among all the techniques used.

The libsvm library is used to implement the SVM classifier for the dataset. The libsvm uses the one vs one classification for all the pairs possible to generate the multiclass classifier. Octave is used to implement the libsvm classifier. We used python script grid.py, provided with the library to find the optimal parameters for the radial kernel to train the data. The parameters for which the validation set was lowest were used to create the final model.

**3.3 Naïve Bayes**

The e1071 library of R is used to implement the classifier for Naïve Bayes. It computes the conditional posterior probabilities of a categorical class variable given independent predictor variables using the Bayes rule. It distributes each class variable as the Gaussian distribution and the parameter tuning is done by the library itself.

**3.3 Neural Networks**

Deep learning is one of the forefront runners in the field of machine learning. The neural network implementation was done using both library and self-written code. But opposite to the initial expectations both the library and ours implementation performed poorly for the data set. We ran various validations and parameter tuning runs but results were not similar to those of the SVM and k nearest neighbors.

**3.4 Weighted Majority Algorithm**

Weighted Majority Algorithm (WMA) is a meta-learning algorithm used to build a joint prediction model from a list of prediction algorithms, which could be any type of learning algorithms, classifiers, or even real human experts. The algorithm assumes that we have no prior knowledge about the accuracy of the algorithms in the list, but there are sufficient reasons to believe that one or more will perform well\*.

We use all but neural network model to build new classifier. We created a script to find the different optimum weights for K-near neighbors, SVM and Naïve Bayes. It runs for all possible weight combinations with step size of 0.01 with range 0 to 100 for each model.

**4 Experiments**

**5 Conclusion**

**6 Future Work**

The work done by us could be extended on various dimensions. The dataset used by us was very small to be able to work properly for the neural network model. The training data could be increased by two methods, (a) collect more data or (b) split the single audio files into multiple files. In our analysis collecting more audio files would be a better approach since reducing the length of audio for training data could reduce the accuracy. Since there might be parts of the songs which may be of different genre than the song as a whole. The feature set used by us is of length 124, this could also be increased. One can calculate these features for fixed intervals of the audio and collate then as new feature set.

Other classification algorithms such as decision trees and random forrest could also be created and added to the WMA algorithm.

**4.1 Citations within the text**

Citations within the text should be numbered consecutively. The corresponding number is to appear enclosed in square brackets, such as [1] or [2]-[5]. The corresponding references are to be listed in the same order at the end of the paper, in the **References** section. (Note: the standard BibTeX style unsrt produces this.) As to the format of the references themselves, any standard reference style is acceptable, as long as it is used consistently.

As submission is double blind, refer to your own published work in the third person. That is, use "In the previous work of Jones et al. [4]", not "In our previous work [4]". If you cite your other papers that are not widely available (e.g. a journal paper under review), use anonymous author names in the citation, e.g. an author of the form "A.Anonymous".

**4.2 Footnotes**

Indicate footnotes with a number in the text. Place the footnotes at the bottom of the page on which they appear. Precede the footnote with a horizontal rule of 2 inches (12 picas).

**4.3 Figures**

All artwork must be neat, clean, and legible. Lines should be dark enough for purposes of reproduction; artwork should not be hand drawn. The figure number and caption always appear after the figure. Place one line space before the figure caption, and one line space after the figure. The figure caption is lower case (except for first word and proper nouns); figures are numbered consecutively.

Make sure the figure caption does not get separated from the figure. Leave sufficient space to avoid splitting the figure and figure caption.

You may use color figures. However, it is best for the figure captions and the paper body to0020make sense if the paper is printed either in black/white or in color.

|  |
| --- |
|  |

Figure 1: Sample Figure Caption

**4.4 Tables**

All tables must be centered, neat, clean and legible. Do not use hand drawn tables. The table number and title always appear before the table. See Table 1.

Place one line space before the table title, one line space after the table title, and one line space after the table. The table title must be lower case (except for first word and proper nouns); tables are numbered consecutively.

Table 1: Sample table title

|  |  |
| --- | --- |
| **Part**  **Description** |  |
| Dendrite | Input terminal |
| Axon | Output terminal |
| Soma | Cell Body (contains cell nucleus) |

**5 Final instructions**

Do not change any aspects of the formatting parameters in the style files. In particular, do not modify the width or length of the rectangle that the text should fit into, and do not change font sizes (except perhaps in the **References** section; see below). Please note that pages should be numbered.

**6 Preparing PostScript or PDF files**

Please prepare PostScript or PDF files with paper size “US Letter,” and not, for example, “A4.” The -t letter option on dvips will produce US Letter files.

Fonts were the main cause of problems in the past years. Your PDF file must only contain Type 1 or Embedded TrueType fonts. Here are a few instructions to achieve this.

* You can check which fonts a PDF files uses. In Acrobat Reader, select menu Files>Document Properties>Fonts and select Show All Fonts. You can also use the program pdffonts which comes with xpdf and is available out-of-the-box on most Linux machines.
* The IEEE has recommendations for generating PDF files whose fonts are also acceptable for NIPS. Please see http://www.emfield.org/icuwb2010/downloads/IEEE-PDF-SpecV32.pdf
* LaTeX users:
  + Consider directly generating PDF files using pdflatex (especially if you are a MiKTeX user). PDF figures must be substituted for EPS figures, however.
  + Otherwise, please generate your PostScript and PDF files with the following commands:
  + dvips mypaper.dvi -t letter -Ppdf -G0 -o mypaper.ps
  + ps2pdf mypaper.ps mypaper.pdf
  + Check that the PDF files only contains Type 1 fonts.
* xfig “patterned” shapes are implemented with bitmap fonts. Use “solid” shapes instead.
* The \bbold package almost always uses bitmap fonts. You can try the equivalent AMS Fonts with command
  + \usepackage[psamsfonts]{amssymb}
  + or use the following workaround for reals, natural and complex:
  + \newcommand{\RR}{I\!\!R} %real numbers
  + \newcommand{\Nat}{I\!\!N} %natural numbers
  + \newcommand{\CC}{I\!\!\!\!C} %complex numbers
* Sometimes the problematic fonts are used in figures included in LaTeX files. The ghostscript program eps2eps is the simplest way to clean such figures. For black and white figures, slightly better results can be achieved with program potrace.
* MSWord 2007 and Windows users (via PDF file):
  + Install the Microsoft Save as PDF Office 2007 Add-in from
  + http://www.microsoft.com/downloads/details.aspx?displaylang=en&familyid=4d951911-3e7e-4ae6-b059-a2e79ed87041
  + Select "Save or Publish to PDF" from the Office or File menu
* MSWord and Mac OS X users (via PDF file):
  + From the print menu, click the PDF drop-down box, and select "Save as PDF…"
* MSWord and Windows users (via PS file):
  + To create a new printer on your computer, install the AdobePS printer driver and the Adobe PostScript Printer Description (PPD) file from
  + <http://www.adobe.com/support/downloads/detail.jsp?ftpID=204>
  + *Note:* You must reboot your PC after installing the AdobePS driver for it to take effect.
  + To produce the ps file, select "Print" from the MS app, choose the installed AdobePS printer, click on "Properties", click on "Advanced."
  + Set “TrueType Font” to be “Download as Softfont”
  + Open the “PostScript Options” folder
  + Select “PostScript Output Option” to be “Optimize for Portability”
  + Select “TrueType Font Download Option” to be “Outline”
  + Select “Send PostScript Error Handler” to be “No”
  + Click “OK” three times, print your file.
  + Now, use Adobe Acrobat Distiller or ps2pdf to create a PDF file from the PS file. In Acrobat, check the option “Embed all fonts” if applicable.

If your file contains Type 3 fonts or non embedded TrueType fonts, we will ask you to fix it.

**6.1 Margins in LaTeX**

Most of the margin problems come from figures positioned by hand using \special or other commands. We suggest using the command \includegraphics from the graphicx package. Always specify the figure width as a multiple of the line width as in the example below

\usepackage[dvips]{graphicx} ...

\includegraphics[width=0.8\linewidth]{myfile.eps}

or

\usepackage[pdftex]{graphicx} ...

\includegraphics[width=0.8\linewidth]{myfile.pdf}

for .pdf graphics. See section 4.4 in the graphics bundle documentation (http://www.ctan.org/texarchive/macros/latex/required/graphics/grfguide.ps)

A number of width problems arise when LaTeX cannot properly hyphenate a line. Please give LaTeX hyphenation hints using the \- command.

**Acknowledgments**

Use unnumbered third level headings for the acknowledgments. All acknowledgements go at the end of the paper. Do not include acknowledgements in the anonymized submission, only in the final paper.

**References**

References follow the acknowledgments. Use unnumbered third level heading for the references. Any choice of citation style is acceptable as long as you are consistent. It is permissible to reduce the font size to ‘small’ (9-point) when listing the references. **Remember that this year you can use a ninth page as long as it contains *only* cited references.**

[1] Alexander, J.A. & Mozer, M.C. (1995) Template-based algorithms for connectionist rule extraction. In G. Tesauro, D. S. Touretzky and T.K. Leen (eds.), *Advances in Neural Information Processing Systems 7*, pp. 609-616. Cambridge, MA: MIT Press.

[2] Bower, J.M. & Beeman, D. (1995) *The Book of GENESIS: Exploring Realistic Neural Models with the GEneral NEural SImulation System*. New York: TELOS/Springer-Verlag.

[3] Hasselmo, M.E., Schnell, E. & Barkai, E. (1995) Dynamics of learning and recall at excitatory recurrent synapses and cholinergic modulation in rat hiippocampal region CA3. *Journal of Neuroscience* **15**(7):5249-5262.