Deep Neural Network Based Poetic Meter Classification Using Musical Texture Feature Fusion

Rajeev Rajan
Dept. of Electronics and Commn. Engg.
College of Engineering,Trivandrum
Kerala,India
rajeev@cet.ac.in

Anu Alphonsa Raju
Dept. of Electronics and Communication Engg.
Rajiv Gandhi Institute of Technology,Kottayam
Kerala,India
anualphonsaraju@gmail.com

***Abstract*—In this paper, a meter classification scheme is proposed using musical texture features (MTF) with a deep neural network (DNN) and a hybrid Gaussian mixture model-deep neural network(GMM-DNN) framework. The performance of the proposed system is evaluated using a newly created poetic corpus in Malayalam, one of the prominent languages in India and compared the performance with support vector machine (SVM) classifier. Initially, a baseline-mel-frequency cepstral coefficient (MFCC) based experiment is performed. Later, the MTF are fused with MFCC. Whilst the MFCC system reports an overall accuracy of 78.33%, the fused system reports an accuracy of 86.66% in the hybrid GMM-DNN framework. The overall accuracies obtained for DNN and GMM-DNN are 85.83%, and 86.66%, respectively. The architectural choice of DNN based classifier using GMM derived features on the feature fusion paradigm showed improvement in the performance. The proposed system shows the promise of deep learning methodologies and the effectiveness of MTF in recognizing meters from recited poems.**

***Index Terms*—meter, timbre, melodic, fusion, rhythm, hybrid, deep learning**

I. Introduction

Music information retrieval (MIR) is a growing field of research with lots of real-world applications and is applied well in categorizing, manipulating and synthesizing music. MIR mainly focuses on the understanding of music data through research, development, and the application of computational approaches and tools. In this paper, the MIR task of poetic meter classification in Indian poetry (Malayalam poems) is addressed using acoustic cues.

Poems are written using the music elements such as rhythm, meter, sounds and imagery at sonic, typographical, sensory and ideational levels [1]. Out of these, the term meter, the periodic arrangements of sequences of stressed and unstressed syllables imposes a systematic regularity in rhythm. In the Western poetic tradition, meters are customarily grouped according to a characteristic metrical foot and the number of feet per line [2]. Each of these types of feet has a certain feel whether alone or in combination. Similarly, the Indian poetry also follows a set of rules in the form of *vrutha* (the synonym for meter) and *prasas*, (meaning, the repetition of sounds). The *vrutha* defines a set of structures to compose the lines of a poem [3, 4], which imposes a unique chanting pattern or musically aesthetic melodic pattern when reading out [5]. In Indian poetic tradition, each syllable of a word is classified as either a laghu (short syllable, U) or a guru (long syllable, '$-$') [3, 6, 4]. The rules are then formulated using ordered sequences (called, *ganam*) of three units in groups of 1, 2 or 4 lines. Since there are two weight distributions for each syllable as either laghu or guru, eight sequences can be formed in a tri-syllabic structure (ref-Table 1). For instance, two lines of a poem written in $K\bar{a}kli$ *vrutha* is shown in Figure 1. We can observe that 8 *ganam* sequences are distributed in two lines. In addition, as a rule in the $K\bar{a}kli$ *vrutha*, each *ganam* consists of two long syllables and one short syllable.

TABLE I
Tri-syllabic feet structure. 'U' and '–' represent short and long syllable respectively

Sl. No	*ganam*	Foot	Mnemonics
1	ya-*ganam*	U – –	ya
2	ra-*ganam*	– U –	ra
3	tha-*ganam*	– – U	tha
4	bha-*ganam*	– U U	bha
5	ja-*ganam*	U – U	ja
6	sa-*ganam*	U U –	sa
7	ma-*ganam*	– – –	ma
8	na-*ganam*	U U U	na

The automatic meter classification has a wide range of application in the generation and the translation of poetry [7]. Machine translation of poetry, probably one of the hardest possible tasks can be addressed using metrical analysis [8]. The systematic archival of a corpus and its aesthetic and emotional perception study can also be done using automatic meteric analysis [9]. Moreover, in this era of enormous web resources, computational tools are inevitable for automatic analysis of poems and sonic devices. The motivation of the proposed work is two-fold. First, to exploit the advantage of combining generative and discriminative models for the novel poetic meter classification task. Second, to show the scope of MTF in poetic meter classification.

978-9-0827-9703-9/19/$31.00 ©2019 IEEE

Fig. 1. Two lines of a poem written in *Kākli vrutha*. Sequences of laghu and guru of are also shown.

A. Related Work

Although there has been significant work in rhythm estimation, there has been relatively little work for the development of specifically designed poetic meter estimation strategies. To the best of our knowledge, most of the previous approaches are based on syllabic structures than the acoustic cues extracted from the poems. In [10], a system is proposed for Persian poetry in which the segmented syllables are classified into long/short syllable based on the features like zero-crossing rate, PARCOR coefficients and the temporal duration [10]. In [1], five different approaches are proposed for poem classification based on the poetic features. In our previous attempt [4], we utilized a limited set of acoustic features with deep neural network architecture to obtain significant results in poetic meter classification. An important aspect of the extended approach is use of hybrid GMM-DNN architecture to learn the metrical traits of recited poems through MTF.

The rest of the paper is organized as follows: Section II describes the proposed system. The performance evaluation is discussed in Section III. The analysis of results is given in Section IV. Finally, the paper is concluded in Section V.

II. PROPOSED SYSTEM

In the front-end, MFCC and MTF are computed. MTF are considered because of their ability to extract the specific and distinguishing traits of a variety of music styles [11]. In the classification phase, two frameworks are employed, namely, DNN and GMM-DNN. The result is compared with that of a SVM classifier. A combined discriminative/generative formulation is derived in the proposed framework that leverages the complementary strengths of both models. A detailed description is given in the following sections.

A. Feature Extraction

In the feature extraction front-end, five sets of features in addition to MFCCs are computed. The MTF include timbral, dynamic, beat histogram-based, melodic and tonality features. From literature, it appears that this is a novel approach of using musical texture features in poetic classification.

1) MFCC: 13 dim MFCCs are computed using frame-size of 40ms and frame-shift of 10ms. MFCCs are widely employed in numerous perceptually motivated audio classification tasks [12], despite of their widespread use as predictors of perceived similarity of timbre [13]. Perceptual filter banks-based cepstral features are based on the computation of a cochleagram, which in some sense try to model the frequency selectivity of the cochlea [11]. The transformations in the MFCC computation crudely approximate the processing in the inner hair cells in the cochlea [13].

2) Timbral and Dynamic features: Timbral features differentiate mixture of sounds that are possibly with the same or similar rhythmic and pitch contents [14]. In timbre space, the perceived (dis)similarity between the sounds is measured and projected to a low-dimensional space where dimensions are assigned with a semantic interpretation such as brightness and temporal variation. Meanwhile, dynamic features model the temporal evolution of the spectral shape over a fixed time duration. Seven specific features, namely spectral centroid, spectral roll-off, spectral flux, zero crossing, low energy, RMS and spectral energy are computed as the part of the experiment [14].

3) Beat histogram features: Beat tracking, the extraction of the rhythmic aspects of the musical content has been a topic of active research in recent years. The rhythmic content in the music signal can be extracted by detecting the salient periodicities of the signal using a beat histogram [15]. In the proposed experiment, seven features, namely, tempo, sum, the salience of the strongest peak, pulse clarity [16, 4], event density, skewness and kurtosis are computed from the beat histogram. To illustrate the importance of beat histogram features, in Figure 2, we display temporal localization of events and beat histograms of two poems written in the same poetic meter, Keka. It can be observed that the similarity in the beat histogram can be used as an important acoustic cue in poetic meter classification.

4) Melodic features and Tonality features : Melodic pitch, which represents the pitch of the leading melody line in the music piece is computed using MELODIA [17]. High level melodic features, namely, pitch deviation, skewness, kurtosis and jitter are computed from the predominant melodic pitch. Whilst there have been studies on meter classification based on prosodic and rhythmic traits, to the best of our knowledge there is no study on meter classification using high-level melodic characteristics. Kernel density estimate (KDE) [18] of melodic pitch of two poems written in the poetic meter, Kavya is shown in Figure 3. Kernel estimators centre a smooth kernel function at each data point to get smooth density estimate as compared to pitch histograms. It is worth noting that KDE of the pitch sequence for the meter *Kavya* shows similarity for both poems in Figure 3. From Figure 4, it appears that the skewness of melodic pitch is an important cue in classification. Apart from melodic features, key and mode are also considered to describe

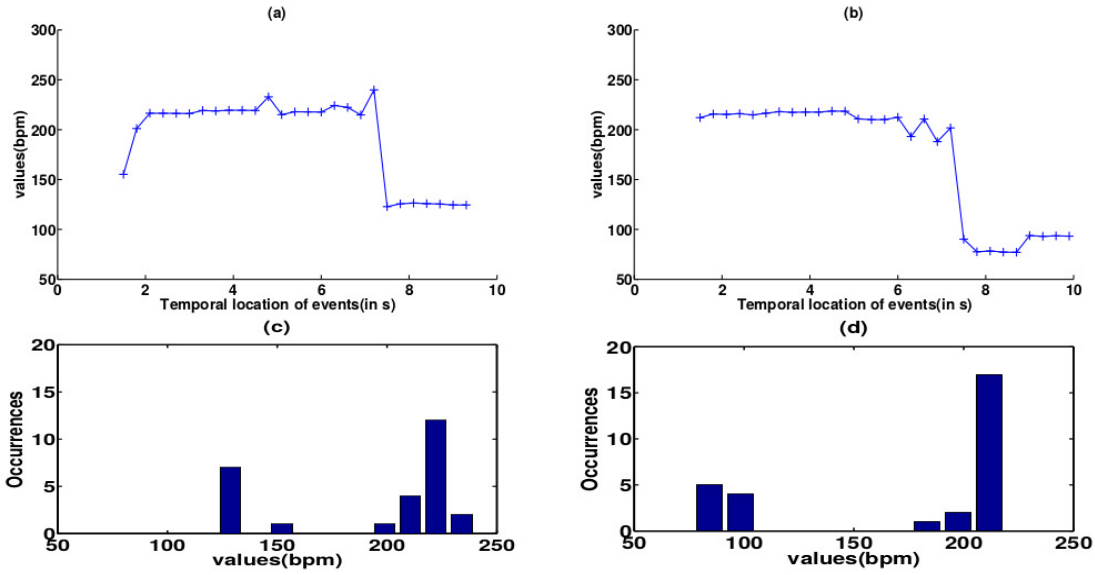

Fig. 2. The temporal location of events (a,b) and beat histogram (c,d) of two poems of Keka meter. Similarity can be seen in temporal location of events and beat histogram

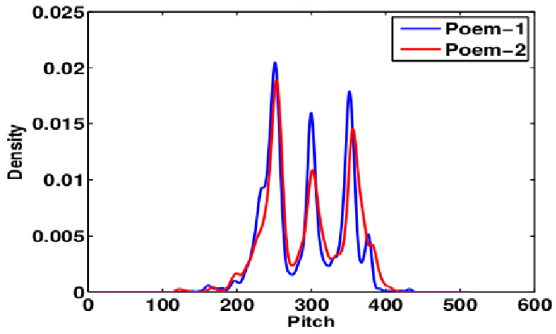

Fig. 3. KDE of melodic pitch for two poems written in the poetic meter Kavya

the tonality structures.

B. Classification Scheme

In the classification phase, SVM, DNN and a hybrid GMM-DNN scheme are employed. In the first phase, baseline-SVM classifier with radial basis function kernel is used. Later, the experiment is carried out using a DNN framework. Our proposed DNN architecture uses three hidden layers (100 nodes per layer) with Adam optimization algorithm [19]. Rectified linear units (ReLUs) have been chosen as the activation function for hidden layers and softmax function for the output layer. In the final phase, a hybrid GMM-DNN framework is employed. 256-mixture GMM, which uses the maximum likelihood (ML) criterion to decide the class identity, is employed in the hybrid experiments. It is a generative model and it fits the training data so that the likelihood of the data given the model is maximized. In contrast, DNN is a discriminative model, and its parameters are trained to minimize the classification error. As a novel attempt, we adopted a hybrid GMM-DNN framework to investigate the promise of log-likelihood features, computed from a GMM, in training a DNN framework. i.e DNN is trained using GMM-log-likelihood of the features, given the meter model.

III. Performance Evaluation

A. Dataset

A database is created in a studio environment for Malayalam, one of the prominent languages in South India. The test dataset consists of six meters, namely Dhruthakakali (Dhruth), Kavyanarthaki (Kavya), Keka (Kek), Nathonnatha (Nath), Omanakuttan (Omana) and Vasanthathilakom (Vasant) with 300 audio tracks, covering all the meters. Out of this, a total of 120 audio files (20 poems/meter) are considered in the testing phase making sure that the same singer does not appear in both training and testing sets. The poems are sung with both male and female singers with background drone, tanpura. Sample audio files, used in the experiment can be accessed at the link.https://mca.rit.ac.in/CASP/downloads.html.

TABLE II
Confusion matrix for MFCC-DNN framework.

Class	Dhruth	Kavya	Kek	Nath	Omana	Vasant
Dhruth	**12**	1	6	0	0	1
Kavya	1	**13**	5	0	0	1
Kek	0	0	**17**	0	2	1
Nath	0	0	0	**20**	0	0
Omana	0	1	9	0	**10**	0
Vasant	0	2	7	0	1	**10**

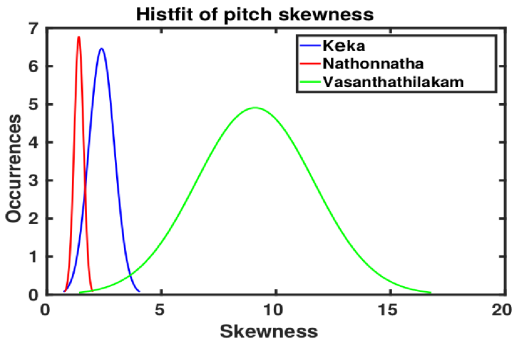

Fig. 4. Histogram of skewness of melodic pitch of poems

TABLE III
CONFUSION MATRIX FOR FEATURE-DNN FRAMEWORK

Class	Dhruth	Kavya	Kek	Nath	Omana	Vasant
Dhruth	**9**	1	0	0	2	8
Kavya	1	**14**	0	0	2	3
Kek	0	0	**20**	0	0	0
Nath	0	0	0	**20**	0	0
Omana	3	0	0	0	**15**	2
Vasant	5	0	2	0	1	**12**

B. Experimental Set-up

In the first stage, experiment is conducted with 13 dim MFCCs. Later, timbral and dynamic features (7 dim), rhythmic features (7 dim) and melodic features (6 dim) are fused in feature level (20 dim) and the experiment is performed. In the final stage, the experiment is extended with the early fusion of MFCC and MTF (13+20 dim). The track level MTF, computed using MIRToolbox [20] are appended to the framewise computed MFCC in all frames. Baseline-SVM and DNN classifiers are implemented using LibSVM and Keras-tensorflow respectively. In DNN experiment, the system was trained for 100 epochs with learning rate of 0.002. In all the phases, around 30 files/meter (apprx. 60 % of the entire dataset) are used for training the models.

IV. RESULTS AND ANALYSIS

The promise of hybrid GMM-DNN architecture and potential of MTF fusion can be analyzed from the results, tabulated in Table VI. From the table, it can be found that for SVM framework, the results reported are 75.0%, 77.5%, 80.83% for MFCC, MTF and its fusion, respectively. For the baseline system itself, MTF played a crucial role in improving the results. It supports our claim that MTF fusion showed improvement due to the complementary information from the individual set. As seen from the results, DNN based experiments resulted in an overall accuracy of 68.33%, 75%, 85.83%, respectively for the above feature sets. It is important to note that fusion in DNN showed significant improvement over SVM. Whilst results of DNN-MTF does not surpass the accuracy reported for SVM, it shows a competitive performance and can potentially be improved with more training data. In the hybrid GMM-DNN system, the experiment shows considerable improvement with 78.33%, 86.86% for MFCC and MTF-fusion, respectively over SVM (improvement of 4% and 6%). It appears that the likelihoods are more effective features to the neural network due to the large dynamic range of the GMM likelihoods citeJoel. Since the MTF are computed in track level, the data was not sufficient to compute generative features for the GMM-DNN framework. It is worth noting that, complementary strengths of the combined discriminative/generative model formulation played a crucial role in improving the accuracy of the hybrid system.

TABLE IV
CONFUSION MATRIX FOR FUSION-DNN FRAMEWORK

Class	Dhruth	Kavya	Kek	Nath	Omana	Vasant
Dhruth	**15**	3	0	0	1	1
Kavya	1	**19**	0	0	0	0
Kek	0	0	**20**	0	0	0
Nath	0	0	0	**20**	0	0
Omana	3	2	0	0	**15**	0
Vasant	3	2	0	0	1	**14**

TABLE V
CONFUSION MATRIX FOR FUSION–GMM+DNN FRAMEWORK

Class	Dhruth	Kavya	Kek	Nath	Omana	Vasant
Dhruth	**17**	0	1	0	2	0
Kavya	0	**19**	0	0	0	1
Kek	0	0	**17**	0	3	0
Nath	0	0	0	**20**	0	0
Omana	0	0	1	0	**19**	0
Vasant	0	0	3	0	5	**12**

In [10], the best match is found out by comparing the sequence of extracted syllables classes with Persian meter styles with an overall accuracy of 69% for 12 classes. But the approach leads to two types of errors: syllabification and classification. Moreover, the syllable classification stage yields substitution, insertion and deletion errors. In our experiment, we rely on the acoustic characteristics instead of syllable-based classification scheme for the proposed task.

The effectiveness of MTF fusion can be well explained using the confusion matrix of the DNN framework given in Tables II,III and IV. From Table II, it can be seen that 80% of the test files are correctly classified for *Kek* and *Nath*, but accuracy is less than 50% for the rest. During the experiment with texture features alone, accuracy for meters *Kek*, *Nath* and *Omana*, becomes more than 75% (Ref.Table III). It is worth noting that MFCC is relatively best to identify *Dhruth* and *Nath* and MTF are best for *Kek*, *Omana* and *Vasant*. From the fusion (MFCC-MTF) results shown in Table IV, it can be seen that individual accuracy becomes greater than 90% for meters, except for *Vasant*. MTF reduced the misclassification errors of *Kavya* and *Omana* to a large extent. The highest classification

TABLE VI
OVERALL ACCURACY FOR ALL THE EXPERIMENTS.

Feature \ Method	SVM	DNN	GMM+DNN
MFCC	75%	68.33%	78.33%
Mus. Texture Feature	77.5%	75%	-
MFCC- Mus. Feature Fusion	**80.83**%	**85.83**%	**86.66**%

accuracy of 100% is reported for *Kek* and *Nath*. By Examining the confusion matrices of the classification results, we found that for the MFCC feature set all the meters confuses primarily with the meter *Kek*. The combination of MTF with MFCC, reduces this confusion, leading to an overall improvement in accuracy. It is quite understandable that rhythmic and melodic features played a crucial role as expected.

The confusion matrix of the fusion experiment for the hybrid GMM-HMM framework is given in Table V. In the hybrid GMM-DNN framework, 26% improvement is observed for *Omana* with an overall accuracy of 86.66%. It is observed that the misclassification error for *Vasant* is still high even in the fusion experiment. A possible cause for this is the piece-wise similarity of '*Vasant*' poems, with other meters while rendering. To end the discussion, from the observation it is clear that acoustic cues have a crucial role in meter estimation.

V. CONCLUSION

Poetic meter classification using the fusion of MFCC and MTF using deep learning methodologies is addressed in this paper. In the first approach MFCC is used, later the experiment is performed with MTF followed by a fusion entire set. The systematic evaluation is done using six meters of Malayalam poetic corpus. The fusion experiment resulted in an overall accuracy of 86.66% in a novel hybrid GMM-DNN framework. The results show the potential of MTF and deep learning methodologies in poetic meter classification task.

REFERENCES

[1] H. R. Tizhoosh, F. Sahba, and R. Dara, "Poetic features for poem recognition: A comparative study," *J. Pattern Reco. Research*, vol. 3, no. 1, pp. 24–39, 2008.

[2] T. Hood, "The rhymester, or the rules of rhyme: A guide to English versification. with a dictionary of rhymes, an examination of classical measures, and comments upon Burlesque, comic verse, and song-writing," *D. Appleton and Company*, 1898.

[3] A. Namboodiri, P. Narayanan, and C. Jawahar, "On using classical poetry structure for Indian language post-processing," *in Proc. of Int. Conf. on Document Analysis and Reco.*, vol. 2, pp. 1238–1242, 2007.

[4] R. Rajan and A. A. Raju, "Poetic meter classification using acoustic cues," *in Proc. International Conference on Signal Processing and Communications (SPCOM)*, 2018.

[5] L. Morgan, R. K. Sharma, and A. Biduck, "Croaking frogs: A guide to Sanskrit metrics and figures of speech," *Createspace Independent Publishing Platform*, 2011.

[6] A. S. Deo, "The metrical organization of classical sanskrit verse," *J. of Linguistics*, vol. 43, no. 1, pp. 63–114, 2007.

[7] E. Greene, T. Bodrumlu, and K. Knight, "Automatic analysis of rhythmic poetry with applications to generation and translation," *in Proc. of the Conf. on Empirical Methods in Natural Lang. Proces.*, vol. 4, no. 10, pp. 524–533, 2010.

[8] D. Genzel, J. Uszkoreit, and F. Och, "Poetic statistical machine translation: Rhyme and meter," *in Proc. of the Conference on Empirical Methods in Natural Lang. Proces.*, pp. 158–166, 2010.

[9] O. Christian, M. Winfried, K. Martin, R. Tim, S. Maren, O. Sascha, and A. K. Sonja, "Aesthetic and emotional effects of meter and rhyme in poetry," *Frontiers in Psychology*, vol. 4, pp. 1–10, 2013.

[10] S. Hamidi, F. Razzazi, and M. P. Ghaemmaghami, "Automatic meter classification in Persian poetries using support vector machines," *in Proc. of IEEE Int. Conf. on Sig. Proces. and Infor. Tech. (ISSPIT)*, pp. 563–567, 2009.

[11] F. Alas, J. C. Socor, and X. Sevillano, "A review of physical and perceptual feature extraction techniques for speech, music and environmental sounds," *Appl. Sci.*, vol. 6, no. 5, pp. 1–44, 2016.

[12] J. Seppanan, "Computational models for musical meter recognition," Masters Thesis, Tampere University of Technology, Department of Information Technology, 2015.

[13] G. Richard, S. Sundaram, and S. Narayanan, "An overview on perceptually motivated audio indexing and classification," *in Proc. of the IEEE*, vol. 101, no. 9, pp. 1939–1954, 2013.

[14] T. Li, M. Ogihara, and Q. Li, "A comparative study on content-based music genre classification," *in Proc.of the 26th Annual Int. ACM Conf. on Research and development in information retrieval*, pp. 282–289, 2003.

[15] G. Tzanetakis and P. Cook, "Musical genre classification of audio signals," *IEEE Tran. on Speech and Audio Proces.*, vol. 10, no. 5, pp. 293–302, 2002.

[16] O. Lartillot, T. Eerola, T. P., and J. Fornari, "Multi-feature modeling of pulse clarity:design, validation, and optimization," *in Proc. of the 9th Int. Conf. on Music Infor. Retri.*, pp. 521–526, 2008.

[17] J. Salamon and E. Gómez, "Melody extraction from polyphonic music signals using pitch contour characteristics," *IEEE Tran. on Audio, Speech, and Lang. Proces.*, vol. 20, no. 6, pp. 1759–1770, 2012.

[18] Y.-C. Chen, "A tutorial on kernel density estimation and recent advances," *Biostatistics and Epidemiology*, vol. 1, no. 1, pp. 161–187, 2017.

[19] G. Dahl, "Deep learning approaches to problems in speech recognition, computational chemistry, and natural language text processing," PhD Dissertation, University of Toronto, Department of Computer Science, 2015.

[20] O. Lartillot, P. Toiviainen, and T. Eerola, "A matlab toolbox for music information retrieval," *in Preisach C., Burkhardt H., Schmidt-Thieme L., Decker R. (eds) Data Analysis, Machine Learning and Applications. Studies in Classification, Data Analysis, and Knowledge Organization. Springer, Berlin, Heidelberg*, pp. 261–268, 2008.