Lyrics based genre classification in Bundeli folk songs

Ayushi Pandey *and Indranil Dutta†

Department of Compuational Linguistics

The English and Foreign Languages University

Hyderabad

December 17, 2013

Abstract

Genre classification is one of the primary tasks of music information retrieval systems. Within the Indian context most work on genre classification has relied on feature extraction from audio input and hardly any work has gone into lyrics based classification [Kini et al., 2011]; [Jothilakshmi and Kathiresan, 2012]. In this paper, we analyse three genres in Bundeli folk music Gaari,Rai and Phag. We use an ensemble of probabilistic and statistical techniques to affect a classification of these genres based ngram token frequencies of both stopwords and POS based lyrical feature. In addition, we successfully demonstrate that a set of common stopwords in Bundeli can be used to categorize song genres and a Bundeli news.

*Electronic address: `ayuship.09@gmail.com`; Corresponding author

†Electronic address: `indranil.dutta.id@gmail.com`

1 Introduction

Song genre classification as one of the primary tasks of music information retrieval has been approached from analysis and classification of audio signal features [Kini et al., 2011, Jothilakshmi and Kathiresan, 2012, Tzanetakis and Cook, 2002]; retrieval of lyrics based features [Howard et al., 2011, Mayer et al., 2008] and approaches which use both audio and lyrics based features for genre classification [Mayer and Rauber, 2011, Neumayer and Rauber, 2007]. Within the Indian context most all work on song genre classification has been restricted to audio feature vector extraction and classification. More precisely, various classification techniques such as Gaussian mixture models (GMM), k-nearest neighbour (kNN) classifiers and Support Vector Machine classifiers (SVM) have been employed to classify Hindustani, Carnatic, Ghazal, Folk and Indian western genres [Kini et al., 2011] and north Indian devotional music [Jothilakshmi and Kathiresan, 2012]. Especially, classification of folk genres has not received any attention. In this paper, we follow a lyrics based approach where ngram token frequencies are used to cross classify between broad text such as news and songs, and within category songs are classifed into specific folk genres in Bundeli; namely, Gaari, Rai and Phag. We utilize both token and type features of unigrams, bigrams and trigrams to successfully classify Bundeli folks into the respective genres. In addition, we also present results that suggest the classification between broad text genres such as news and songs can be successfully accomplished using common stopwords in Bundeli. The paper is organized as follows: In section 2, we elaborate on standard lyrics based classification methods, following that, in section 3 we provide details of the materials and methods that have been employed in this paper. In section 4 we present the results from the study where we provide details on broad text classifcation using common stopwords, as well as statistical and probabilistic techniques for within genre classification. Section 5 provides details of our analysis and also future directions for further research in the fledgling area of song genre classification using statistical techniques.

2 Lyrics based classification

One of the most popular lyrics based classification techniques is the Bag of Words technique. Combined with algorithms like Naïve Bayes, J48 and SMO, these help to identify multilingual genres. Also, there are various ensemble approaches that give the accuracy of the best single model trained. This model is then further selected for any paticular feature or dataset. While, all these approaches help categorize between broad genres, within genre classification has not yet received much attention. Lyrics based approaches help by paying attention to the 'content semantics' of the song. In addition, lyrics based approaches allow us to tackle the problem of low sound quality in audio. Features are selected from both audio and lyrics, combined and then a downscaling is performed using different dimensionalities matching the dims of the audio. Since music is a multi-modal system, a multi-modal approach results in better discriminating accuracy[**?**].

In our approach, we use an ensemble of statistical and probabilistic techniques by integrating ngram token frequencies, POS based features and probability densities to categorize Bundlei folk songs into their respective within-genre classification; namely Gaari, Rai and Phag.

We predict that owing to the intense repetition in our three genres, Gaari following a specific pattern, Rai another and Phags none, we will be able to test the within genre differences. Ngram token frequency probability densities will help classify these three sub-genres and account for the lyrical characteristics of these songs. The consideration of stopwords has mostly been used for multilingual classifications. For lyrics classification, generally it is not thought of as the best idea. However, we assert that genres can be identified based on stopword classification. For if a genre uses a certain set of stopwords more than the other, it could be a proabable indicator of the genre. The probability densities of token frequencies of ngrams correspond to the structural pattern of the song. The structural pattern of the written song is important in performing lyrics based classification of genres. The features that we have used are all POS based features. Their position decides their role in the song or speech. These are all grammatical categories of various kinds. ((The stopwords thus derived are used to classify between song and news, and as features to identify between genres.))(not nice) Gaaris are predicted to have more imperatives, since they are marriage songs and follow the most conversational tone among other genres. Vocatives should also feature both in rais and gaaris, since they follow an informal tone ove phag.

3 Materials and methods

In this section, we provide a brief description of the corpora used for classifying Bundeli between songs and news. We also discuss details of corpus creation and annotation. In an effort to approach classification based on ngram counts and token frequencies of the songs, we limit the annotations to a minimal set so as not to overspecify the statistical models, and keep the numbers of parameters responsible for classification to a minimum.

3.1 Ngram token frequencies

Ngrams and ngram counts were generated using modified python scripts that are able to process Unicode (UTF-8) encoding[1]. The output was further modified to generate token frequencies from the raw ngram counts for each genre using the following equation:

$$TokenFrequency_{Genre} = \frac{NgramCount}{TotalNgrams} \tag{1}$$

[1]http://jaganadhg.freeflux.net/blog/archive/2009/07/22/n-gram-library-for-indian-languages.html

This way, the ngram counts are scaled by the total number of ngrams in each genre and help compare lyrics data from variable number of songs. In our corpus of songs, we find Gaari and Rai represented by 4 songs each, and Phag by 18 songs. Scaling the ngram counts also allows us to compare between the probability densities which provide insight into the lyrical structure of the various genres.

3.2 News corpus: Creation and annotation

To establish a standardisation for our corpora, we created a news corpus from an online publication of a newspaper[2]. Carefully selected news articles from the Mahobia region was included into the corpus, to reduce any dialectal variation and consequent amibguity. The dialect from the Mahobia region complied best with the songs in our corpus. The annotation schema used in the corpus was the $< EOS >$, which marks the end of the sentence.

3.3 Song corpus: Creation and annotation

The songs come from various sources. Our first source was collection from an oral tradition of singing from regions near Sagar, Madhya Pradesh. Regions as Tikamgarh and Orchha and Jhansi also recognise this tradition, and online videos from a Jhansi based casette company, Kanhaiya Casette became our second source. The Phags, however, were mined from a web resource for poetry and prose which contains collections of the famous Bundeli poet Isuri [3].

We annotated the songs in each genre using the following tags; the beginning of verse $< BOV >$, the end of verse $< EOV >$ in phags. Considering the intense repetition of Rais (owing to their dance-form structure), we added tags for repeated lines¡RL¿ and ends of repeated lines.<ERL>. The call in the beginning of the Rai, which uses a long aalap

[2] http://www.khabarlahariya.org/?cat=64

[3] http://www.kavitakosh.org

for "Limtera" following little internal structure was tagged $< LIM >$ and $< ELIM >$ at its beginning and end. Here too, we had $< BOV >$ and $< EOV >$ for wherever the song required it. Owing to its conversational nature of repetition and inclusion, the Gaari annotation scehema is also replete with the $< RL >$ and $< ERL >$ Gaaris and Rais distinguish another special feature, the chorus. In the Gaaris, the leader sings the chorus and the rest of the party follows it. Whereas in Rais, the chorus is sung to repeat the dance move. Hence, we use $< C >$ and $< EC >$. Between the end of verse and chorus, there is a chunk from the chorus that rhymes with the last line of the verse. This repeated chorus chunk, our annotation schema recognises as the repeated chorus line, $< RCL >$ and $< ERCL >$.

3.4 Lyrics based feature matrix

To further identify differences between songs, we decided to look at individual features in each genre and also our training corpus. Hence, for each subset of the corpus, we created lists of the unigrams based on the following features, stopwords (which included but was not limited to common function words) and part-of-speech (POS) based features such as personal pronouns, wh-words, vocatives and imperatives. Feature vectors for unigram token frequencies for both stopwords and POS based features were also created for broad text and genre classification.

4 Results

4.1 News and song corpus classification

Most common text classification methods, including music genre classification, begin by removing the common stopwords derived from a generalized corpus. Common stopwords include function words, personal and relative pronouns and most high frequency unigrams.

Token frequencies of a set of 7 common stopwords was used to One way ANOVA with Broad Genre as predictor shows a significant main effect $F[1,26]=5.438; p<0.05$. As the boxplot in Figure 1 shows, stopword token frequencies are significantly higher in news compared to songs. Thus, stopwords when included in any corpus, can help classify news from songs, even though more commonly stopwords are excluded during the preprocessing stages of classification.

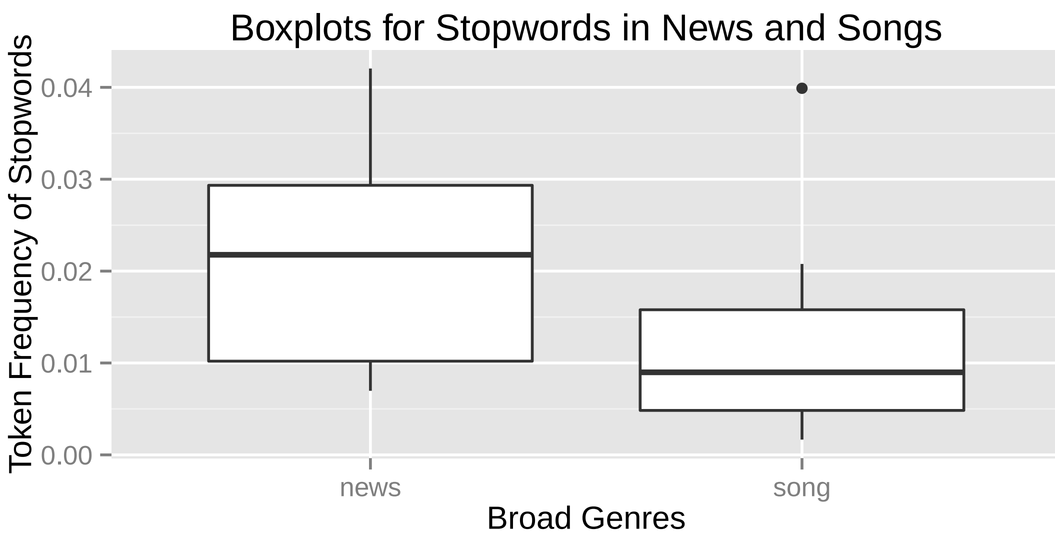

Figure 1: Higher stopword token frequencies in the Bundeli news corpus compared to the songs

4.2 Kernel Density Estimates

Kernel Density Estimates (KDE) of ngram terms are generated using Equation 2 below. KDE is a non-parametric estimate of the probability density function of random variables, in this case the counts of the ngrams. KDEs allow for better inferences about the population, based on a real and finite data sample. KDE makes it possible for us to closely examine the probability density function of ngrams for the various genres. Based on the KDEs we make inferences about the possible presence of categorical features in the sample based on smoothed bins and probability density. Equation 2 shows the KD function estimator. In our estimates, we use a normal kernel such that K(x) = ϕ(x), where ϕ is the standard normal density function.

$$\hat{f}_h(x) = \frac{1}{n}\sum_{i=1}^{n} K_h(x - x_i) = \frac{1}{nh}\sum_{i=1}^{n} K\frac{(x - x_i)}{h} \tag{2}$$

We used R Studio to generate KDEs to identify unigrams, bigrams and trigrams for Gaaris and Phags and Rais. To see a fair distribution across genres in terms of token frequencies, we created graphs between different ranges. The different ranges showed windows of the distribution, enabling us to make the following three-way generalisations. The gaaris show a high concentration of trigrams, bigrams and unigrams in the lower limit of the window lowest 2- and the highest 5 counts. The phags have peaks of their bigram and trigram density concentrated around the 2 and 3 counts. Rais, however, have their ngram density distributed such that one can see some concentration towards lower counts, but it tapers down towards the higher ones. Each line of the verse is repeated in the Rai, this could give rise to such a distribution. So each repeated line, every line repeated even once falls in this categorisation.

The next distribution was between the lowest 5- and the highest 10. Here, we saw a wide difference between the three genres. The unigrams for gaaris have a large concentratiof unigrams correspnding to the corresponding token frequency of 11 and 12 counts. For Phags, the trigrams disappear in this window. Because of little or no repetition in phags, we can justify this. Rais, spread in the corresponding count frequency of 9-12 counts, and then narrow down, showing lesser frequency. This may look like an uninteresting distribution for Rais, but this is actually the largest window w.r.t Rais. Repeated lines of the verse account for this recurrence in Rai.

The third window shows probability density distribution such that trigrams in Gaaris shoot up in concentration of token frequencies corresponding to 16, clearly indicating those coming from the repeated chorus. Bigrams also show considerable bumps in (chorus repeats 4 times). Looking at phags, the bigrams and trigrams vanish completely from the distribution. Because of little or no repetition, phags cannot show such a high bigram or trigram concentration. The new elements brought in by the phag have already been exhausted in the first two graphs. For Rais, this is the window that shows the repeated chorus line. These vary on the number of verses the song has, hence there are no exact peaks. Finally, gaaris correpsonding to the 16 count reach their peak in the last window. This happens both in the case of bi and trigrams. ((The graphs might show two different pictures, but actually that is only because of the overlapping lower-higher limits.)) Once again, this is because of chorus repetition and the repeated lines. Bigrams are more frequent than trigrams because of POStagging. For Phags, bigrams and trigrams are completely off the scene in this last graph. Here, finally, the Rais show their peaks in the lower limit of the x axis, corresponding to 16-19 count frequencies. The peaks in the Rai are primarily due to the repeated lines and the choruses. Whenever there is a defined chorus, the repeated pattern behaves in this way. Since there is no obligation for a chorus, this is also a small window for Rai.

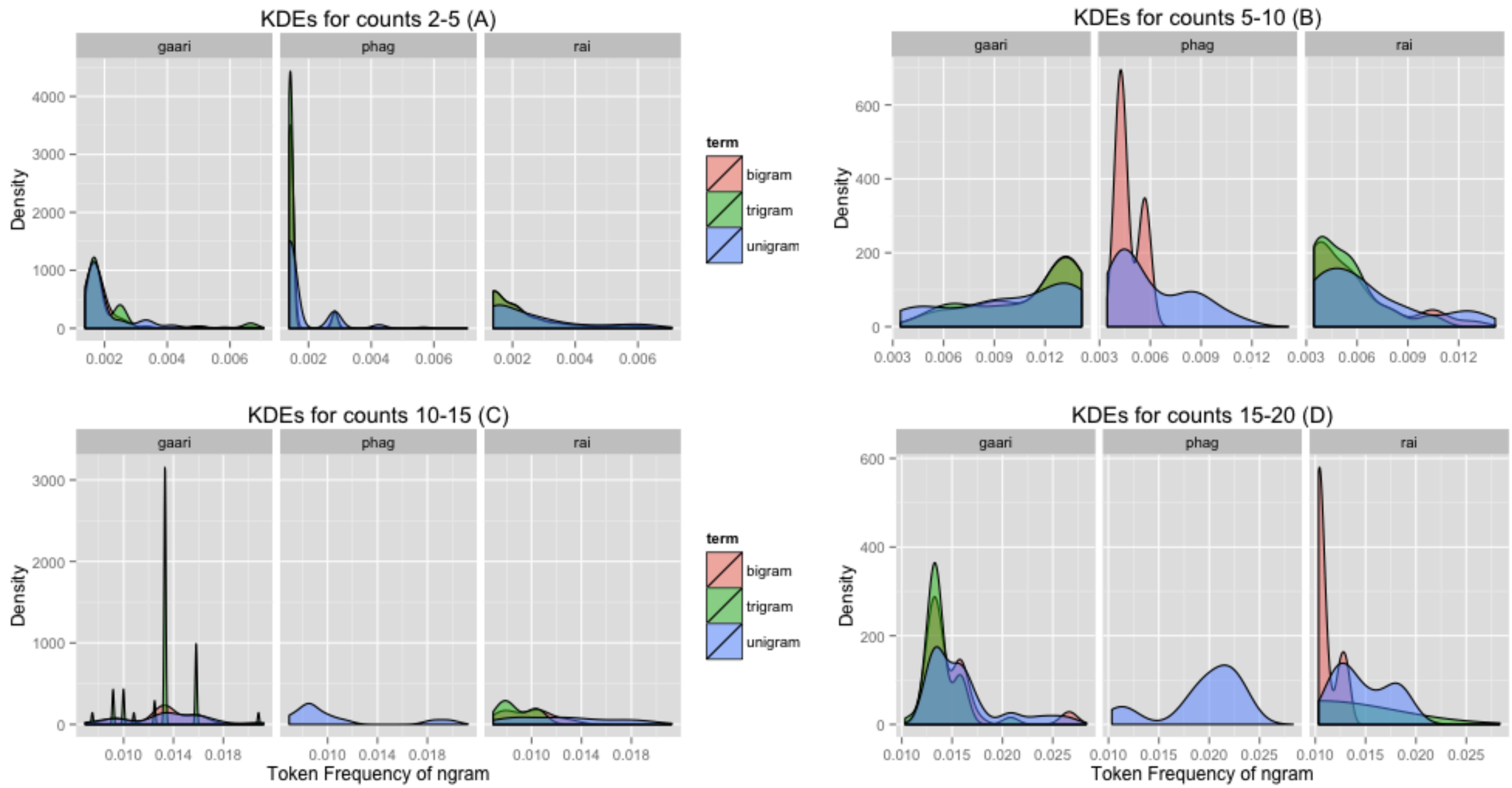

Figure 2: KDEs

4.3 Song corpus classifcation

While the stopword based classification produces satisfactory categorization between broad text news and songs, for classifying the three genres; Gaari, Rai and Phag, we utilize unigram feature vectors of POS based features. These include imperative forms, wh-terms and personal pronouns. We note that the imperative forms allow for a clear two way distinction between the genres. A one way ANOVA with token frequency as the dependent variable and genre as the independent variable reveals a main effect of genre on imperative use; $F[2,11]=6; p<0.05$. A post-hoc TukeyHSD comparison reveals that imperative use allows for a distinction between Gaari and Rai, and Gaari and Phag, however, the distinction between Rai and Phag in not significant. The order of imperative use is such that *Gaari>Rai=Phag*. This distinction can be seen in the boxplots in Figure 2 below.

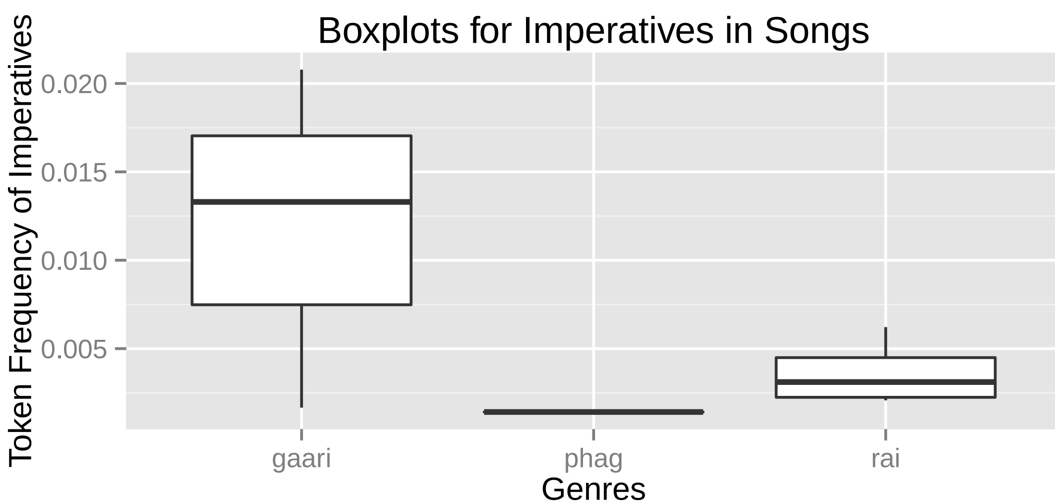

Figure 3: Token frequencies for imperative use in Gaari, Phag and Rai

5 Analysis and conclusion

References

Sam Howard, Carlos N. Silla Jr., and Colin G. Johnson. Automatic lyrics-based music genre classification in a multilingual setting. In *Thirteenth Brazilian Symposium on Computer Music*, 31st August–3rd September 2011 2011.

S. Jothilakshmi and N. Kathiresan. Automatic music genre classification for indian music. In *International Conference on Software and Computer Applications (ICSCA 2012)*, 2012.

S. Kini, S. Gulati, and P. Rao. Automatic genre classification of north indian devotional music. In *Proceedings of the National Conference on Communications (NCC)*, pages 1–5, Jan 2011, Bangalore, India 2011. doi: 10.1109/NCC.2011.5734697.

Rudolf Mayer and Andreas Rauber. Music genre classification by ensembles of audio and lyrics features. In *Proceedings of the 12th International Society for Music Information Retrieval Conference*, pages 675–680, Miami (Florida), USA, October 24-28 2011. `http://ismir2011.ismir.net/papers/PS6-4.pdf`.

Rudolf Mayer, Robert Neumayer, and Andreas Rauber. Rhyme and style features for musical genre classification by song lyrics. In *In Proceedings of the 9th International Conference on Music Information Retrieval (ISMIR'08*, 2008.

Robert Neumayer and Andreas Rauber. Integration of text and audio features for genre classification in music information retrieval (accepted for publication. In *Proceedings of the 29th European Conference on Information Retrieval (ECIR'07*, pages 724–727, 2007.

G. Tzanetakis and P. Cook. Musical genre classification of audio signals. *Speech and*

Audio Processing, IEEE Transactions on, 10(5):293–302, 2002. ISSN 1063-6676. doi: 10.1109/TSA.2002.800560.