Bundeli Folk-Song Genre Classification with kNN and SVM

Ayushi Pandey, Indranil Dutta
ayuship.09@gmail.com, indranil@efluniversity.ac.in

Folk song genre classification

Purpose: Folk song within-genre classification using kNN and SVM

Data: Bundeli folk songs from three genres; **Gaari**, **Rai** and **Phag**

Motivation

- Song genre classification as one of the primary tasks of music information retrieval approached from analysis and classification of audio signal features [1, 2, 3]
- Retrieval of lyrics based features [4, 5]
- And approaches which use both audio and lyrics based features for genre classification [6, 7]
- Western context - Stochastic and probabilistic methods used for popular, classical, and folk music genre classification; [8]
- Indian context - Only audio feature vector extraction and classification between Hindustani, Carnatic, Ghazal, Folk and Indian-Western genres [1] and north Indian devotional music [2]; Gaussian Mixture Models (GMM), k-nearest neighbour (kNN) and Support Vector Machines (SVM)
- Lyrics based classification in the context of Bollywood music to identify specific features of Bollywood song lyrics [9].
- Classification of **folk genres** has not received any attention so far

♪ Our work: Two-fold approach

1. Classification between broad text genres such as news and songs using common stopwords in Bundeli
2. 'Big data' based machine learning approaches to classify a small corpus of Bundeli folk songs into specific genres; Gaari, Rai and Phag

Corpus

Song corpus Collection from an oral tradition of singing from regions near Sagar, Madhya Pradesh

- Online video and audio cassettes
- Phags were web-mined from a collection of Bundeli poet **Isuri** [kavitakosh.org]
- Lyrics were orthographically hand-transcribed

News corpus Articles from Mahoba region web-mined [10]

Table 1: Details of the news and song corpora

Type of text	Number
News Articles	98
Gaaris (G)	37
Rais (R)	39
Phags (P)	40
Total songs (G+R+P)	116

kNN Classification

- kNN classification yields an accuracy of 86.95% when k=2 nearest neighbours are used to measure the distance between the *tfidf* scores with a precision of 0.818 and an F-score of 0.9
- Following a 10-fold cross-validation and the lowest test, CV mean and CV Standard errors an optimal K=2 is found to be the best kNN classifier for our dataset

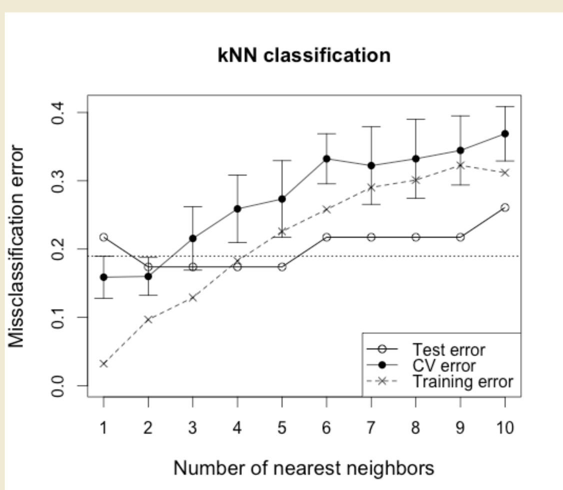

Figure 3: Misclassification errors from kNN training, testing and cross-validation

Structural Description using Kernel Density Estimates

Kernel Density Estimates (KDEs) of ngram terms are a non-parametric estimate of the probability density function of counts of the ngrams. 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 1 shows the KD function estimator. We use a normal kernel such that K(x) = $\phi(x)$, where ϕ is the standard normal density function.

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

Unigrams and bigrams Unigrams and bigrams show a high degree of overlap between the three genres
Trigrams The peaks of the distribution differ, the probability density is extremely low. These terms would be eliminated in the sparsity calculation and cannot be termed as predictors of the genre-variation.

The KDEs establish the existing overlap in the three genres

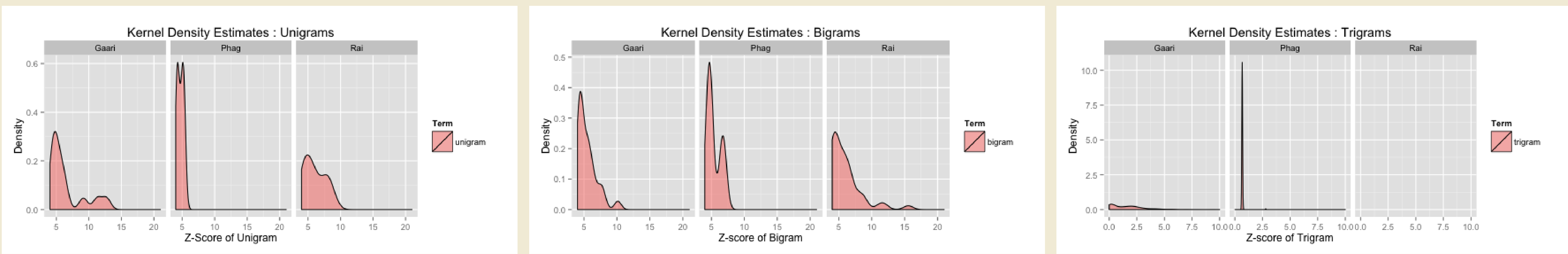

Figure 1: Kernel Density Estimates for Unigrams, Bigrams and Trigrams

News and song separation

- Common classification methods begin by removing stopwords derived from a generalized corpus
- We use the token frequencies of a set of 7 stopwords to separate news from songs
- One way ANOVA with Broad Genre as predictor shows a significant main effect $F[1,26]=5.438; p<0.05$
- Stopword token frequencies are significantly higher in news as compared to songs
- News corpus exhibits a type-token ratio of 0.028%
- Song corpus exhibits a type-token ratio of 0.014%; nearly half the lexical diversity of the news corpus

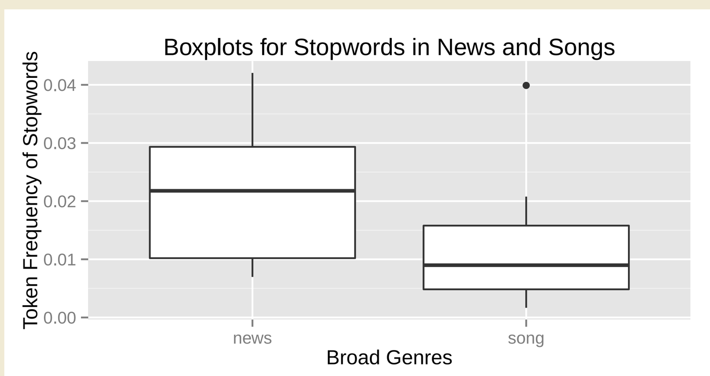

Figure 2: Higher stopword token frequencies in the Bundeli news corpus compared to the songs

SVM Classification

- The e1071 package in R is used to perform a SVM classification on the songs dataset [11]. The term-document matrix used for kNN classification is passed to the SVM algorithm
- An 80-20% partition is used for training and test, respectively.
- The training dataset contains 93 songs, and the test dataset has 23 songs.
- A 5-fold cross-validated SVM exhibits a diagonal accuracy of 91.3% and kappa accuracy of 85.7%
- Table 2 shows the confusion matrix as predicted by the SVM Classifier.

Table 2: True positives and false positives for SVM Classification

	True		
Predictions	Gaari	Rai	Phag
Gaari	11	2	0
Rai	0	3	0
Phag	0	0	7

Preprocessing and methods

Following collection and compilation of the corpus

- Stopwords, punctuation marks and white-spaces were removed
- Sparsity was set to 85%
- Term-document matrix: Each song file converted to a vector space of term frequency-inverse document frequencies (*tfidf*)
- The *tf* (term frequency) scores calculated with Equation 2 and the *inverse document document frequency (idf)* scores with equation 3

$$tf(t,d) = 0.5 + \frac{0.5 \times f(t,d)}{max\{f(w,d) : w\epsilon d\}} \tag{2}$$

$$idf(t,D) = log\frac{N}{|d\epsilon D : t\epsilon d|} \tag{3}$$

The tf-idf scores are a product of the tf and idf scores from equations 2 and 3.

$$tfidf(t,d,D) = tf(t,d) \times idf(t,D) \tag{4}$$

The product of the tf and idf scores are used to train a classifier with 10-fold cross-validation.

Naïve Bayes Classification

- Same as in SVM classification, we use e1071 package in R to perform a Naïve Bayes classification on the songs dataset [11]
- An 80-20% partition is used for training and test, respectively
- The Naïve Bayes performs classification with a diagonal accuracy of 78.2% and kappa accuracy of 68.4%. Table 3 shows the confusion matrix as predicted by the Naïve Bayes Classifier.

Table 3: True positives and false positives for Naïve Bayes Classification

	True		
Predictions	Gaari	Rai	Phag
Gaari	6	0	0
Rai	5	5	0
Phag	0	0	7

Conclusions

What we find

- Lyrics were separated from standard Bundeli texts by performing a broad-genre classification using stopwords and lexical diversity measures
- Machine-learning techniques to successfully classify the three genres
- Our findings report that popular methods of classification that are employed on 'big data' can be used to perform within-genre classification
- SVM and kNN Classifiers perform better than Naïve Bayes classifier.

References

[1] 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.

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

[3] G. Tzanetakis and P. Cook. Musical genre classification of audio signals. *Speech and Audio Processing, IEEE Transactions on*, 10(5):293–302, 2002.

[4] 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.

[5] 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.

[6] 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.

[7] 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.

[8] Wei Chai and Barry Vercoe. Folk music classification using hidden markov models. In *Proc. of International Conference on Artificial Intelligence*, 2001.

[9] Aseem Behl and Monojit Choudhury. A corpus linguistic study of bollywood song lyrics in the framework of complex network theory. In *International Conference on Natural Language Processing*. Macmillan Publishers, India, 2011.

[10] Khabarlahariya: a product of women, media and news trust. http://www.khabarlahariya.org/?cat=64.

[11] Evgenia Dimitriadou, Kurt Hornik, Friedrich Leisch, David Meyer, , and Andreas Weingessel. *e1071: Misc Functions of the Department of Statistics (e1071), TU Wien*, 2010.