

Review on Evolutionary Change and Expressiveness in Lyrics of Indian Filmy Songs with Natural Language Processing

¹Darshita Pathak ²Dr Tejas P. Patalia³Dr Chirag A. Patel⁴Dr Sohil D. Pandya

¹Research Scholar²Guide³DPC Member⁴DPC Member

^{1, 2,3,4} Gujarat Technological University Ahmedabad Gujarat India

Abstract : Examination in Natural Language Processing is extending in various spaces and is saturating all parts of existence with time. With each progression, the assortment of text that can be prepared is developing. One such area is verses handling. Songs are fundamental to the music and entertainment world and can be broke down to acquire significant data, for example, type, topic, mind-set, and so forth of the tune and supplement the data assembled by the investigation of its sound highlights. Bollywood, the Indian entertainment world makes a ton of income utilizing melodies. The quantity of tunes produced by Indian film industry is gigantic and is a rich wellspring of sound and literary information for Natural Language Processing assignments. It additionally offers us a chance to deal with information in Hindi which is a generally less investigated field.

Index Terms – Lyrics, Music, Hindi Songs, Natural Language Processing, Filmy Songs, Expressiveness, Information Retrieval.

1. Introduction

Music information retrieval (MIR) is the interdisciplinary science of retrieving information from music. MIR is a small but growing field of research with many real-world applications. Those involved in MIR may have a background in musicology, psychology, academic music study, signal processing, machine learning or some combination of these. Music Information Retrieval (MIR) has been defined by Stephen Downie as ‘a multidisciplinary research endeavor that strives to develop innovative content-based searching schemes, novel interfaces, and evolving networked delivery mechanisms in an effort to make the world’s vast store of music accessible to all’. And this MIR mainly deals with Music Content, Music Similarity and Music Psychology areas^[1]

2. Related Work

Music information retrieval systems seem to have reached an impasse. A significant number of attempts at creating efficient and effective systems to search through large databases of music have brought about almost the same number of solutions or approaches. MIR is being used by businesses and academics to categorize, manipulate and even create music^[1].

The inalienable method to put together the tune is to characterize the tunes that this isolating the tunes into grouping of pieces as per the classification portrayed by their music content. Classification of the content negative, positive and nonpartisan was the principal consolation to distinguish feelings in a sentence. The most widely recognized approaches to mark the classes/artist. Moreover, the fuzzy meaning of music classification can yield the multi-order of individual songs. The melodies ought to be classifiable into eh phase of programmed order. The automatic classification of songs are categorized into 5 different stages (i) Audio content format based. (ii)Symbolic features content based. (iii)Lyrical based approach. (iv)Web community data based. (v)Hybrid approaches. Other than this data set model also lays an important roles of classification. So, based on the machine learning Cluster based classification I.e. Unsupervised learning will be used in it. In that cluster based algorithm some of the regressions are studies. A song is composed of both verse and chorus segments.

User’s mood can be drive him in selecting some song, and that melody can be utilized to conjure slants in the audience members, which can make the audience members being joined to the feeling of the tune or it can make audience members change their current mind-set moreover.

2.1 Statistics and Machine Learning

- Computational methods for classification, clustering, and modeling — musical feature extraction for mono- and polyphonic music, similarity and pattern matching, retrieval
- Formal methods and databases — applications of automated music identification and recognition, such as score following, automatic accompaniment, routing and filtering for music and music queries, query languages, standards and other metadata or protocols for music information handling and retrieval, multi-agent systems, distributed search)
- Software for music information retrieval — Semantic Web and musical digital objects, intelligent agents, collaborative software, web-based search and semantic retrieval, query by humming, acoustic fingerprinting

•Music analysis and knowledge representation — automatic summarization, citing, excerpting, downgrading, transformation, formal models of music, digital scores and representations, music indexing and metadata.

2.2 Lyrics

Lyrics are a set of words that make up a song, usually consisting of verses and choruses. The writer of lyrics is a lyricist. The words to an extended musical composition such as an opera. LRC is a computer file format that synchronizes song lyrics with an audio file, such as MP3, Vorbis or MIDI ^[2,3].

When an audio file is played with certain music players on a computer or on modern digital audio players, the song lyrics are displayed. The lyrics file generally has the same name as the audio file, with a different filename extension. ^[4] For example, song.mp3 and song.lrc.

```
[00:12.00]LINE 1 LYRICS
```

```
[00:17.20]LINE 2 LYRICS
```

```
...[MM:SS.XX]LAST LYRICS LINE
```

2.3 Indian Filmy Songs

Hindi is the most broadly communicated in language in India, practically 41% of the Indian populace being its local speakers. It is additionally one of the Indian dialects with a decent collection of devices and assets for Natural Language Processing research. Music from Bollywood comprises practically 71% of music deals in India. This is a tremendous market and it would be basic to chip away at better association of this information. Individuals decide to tune in to various music dependent on their mind-set and circumstances. It is subsequently, of help to clients to have the option to coordinate advanced music libraries as per their opinion or utilize this data in programmed playlist age undertakings as investigated in [5]. Watchword age from verses would help recognize the opinion of the melody and furthermore in different errands, for example, subject discovery [6].

Bollywood tunes have encountered an impressive change as far as verses over the previous many years. Some time in the past, Bollywood melodies contained generally Hindi words, yet these days lyricists use loads of English words to communicate. Till date, we don't know about any deliberate quantitative examination of how it has changed and how the advancement occurred throughout the long term.

Computational examination of Bollywood melodies has not gotten a lot of consideration by analysts. In any case, a few investigations on Music Information Retrieval (MIR), utilization of melodic scales in Bollywood music and symbolism in contemporary verse, computational examination of basque tune assortments, tune highlights in well known tune and corpus phonetic investigation of 50 years of Bob Dylan verses have been completed in the new many years. A lot of work has been done on the music state of mind grouping dependent on sound, verses, and social labels and all together.

Programmed techniques are needed to characterize music as indicated by temperaments. This has been accomplished by building a framework for characterizing mind-sets of Hindi tunes which utilizes sound related highlights, for example, musicality, wood and force [8]. Music Information Retrieval Evaluation eXchange (MIREX) has been utilized during the trial. What's more, the choice tree classifier has been utilized for grouping purposes and a normal exactness of 51.56% has been accomplished [8]. Moreover, the advancement of Bollywood music concerning the utilization of melodic scales has been the theme under examination. Through an investigation, different social impacts on Bollywood music and its arrangers throughout the years were distinguished and measured [7]. From the examination, some striking realities were uncovered about the scale utilization designs that were helpful in figuring intriguing guesses which can be checked genuinely.

3 Evolutionary Change in Lyric Content based Analysis

As grouping a song, exclusively dependent on verses is a difficult undertaking, Ashley M et al. [3] proposed a strategy to distinguish the enthusiastic extremity of tune through verses utilizing Natural Language Processing. A dataset containing 420 melodies had been thought of, which contained equivalent number of tunes from negative and positive feelings. Last.fm API [32] was utilized to recover labels for melodies. Estimation examination of tunes is an intricate errand since tunes may fire up with positive feeling however can wind up in negative feeling and the other way around, correspondingly, a few refrains of melody verses may establish to a specific feeling and furthermore tunes can talk negative feelings about good things and the other way around. To address each one of those issues, various calculations were proposed. Word list is most straightforward calculation that depends on word tallies. The tunes are ordered by keeping a checks of the words that shows up in the verses. Here, the calculation circles through the expressions of the tunes. In word reference, for melody characterization, the calculation circles through the verse words, deciding if each word is available in both positive and the negative word references or in single or in none.

Cosine similarity method based on term frequency-inverse document frequency is used to find similarity between an input song and to the songs in training set. The proposed method could not deal with negations. Results shows that song lyrics when considered alone does not show promising results but it improves the classification accuracy when additional features such as audio are combined to lyrical features. Table.3.1 lists out the features and methods that various authors have used to classify songs.

Author Name	Features/Method	Dataset	No. of Songs	Emotion
Ashley M et al. [12]	TF, IDF	Created manually	420	Positive, Negative
Yanqing Xia et al. [11]	Sentiment related words	5SONGS	2,653 Chinese Pop songs	Positive, Negative
Corona et al. [13]	TF,IDF,BM25	Million Songs	32,302 English songs	--
Shanmugapriya et al. [14]	Hidden Markov Model	Collection of songs from Website	--	Happy,Angry,Sad
Emil Ian et al.[15]	TF, IDF, Key Graph	Collection of songs from Internet	200 Songs	--
Doran Walsten et al.[17]	K-Means Clustering	Collection of songs from Website	--	Classic rock, county, grunge, modern rock, pop, r&b, rap
Govind Sharma et al.[16]	Bag of Words/ Latent Dirchlet Allocation	Collection of songs from Internet	--	Happy, sad, angry, Tired, Love, Funny
Charulatha et al.[18]	Rakshini Algorithm	--	--	Happy, sad, calm

Table 3.1. Features & Methods for Lyrical Analysis

	SVM	NB	SVM+NB
Whole Song	60.60	52.3	70.2
Beginning of a song	67.5	57.3	75.7
Ending of a song	64.4	55.8	72.4

Table 3.2 Sentiment Classification with Lyric Features [10]

Lexical Choices for Women	Lexical Choices for Men
Munni BADNAM	Sing is KING
Chamak Challo	Shehansha BOSS
Dance Basanti	KHILADI Bhaiya
Afghan Jalebi	DABANGG
Fevicol Se	Kukkad Kamal
Khallas	Jatt Ludhiyane Da
Laila Laila	Bahubali Title

Table 3.3. Comparative Analysis of Lexical Choice of Women v/s Men^[19]

Words	Number of occurrences	Percentage Occurrence (%)
baby	110	11.20
love	108	11.0
dance	85	8.66
girl	69	7.03
twist	54	5.50
shake	49	4.99
darling	49	4.99
rock	43	4.38
floor	42	4.28
welcome	41	4.18
party	40	4.07
know	40	4.07
right	39	3.97
all	38	3.87
sunny	36	3.67
feel	34	3.46
now	31	3.16
lucky	29	2.95
touch	23	2.34
naughty	22	2.24

Table 3.4 Top 20 Foreign Words^[10]

4 Expressiveness in Lyrics

Expressives are generally utilized in beautiful articulation and scholarly classifications all through the subcontinent. The Hindi entertainment world is one of the biggest entertainment worlds on the planet, delivering around 1,000 motion pictures a year (Matusitz and Payano 2011).

4.1 Reduplication in Hindi

Reduplication is a typical and beneficial element of all South Asian dialects, paying little mind to language family (Abbi 1985). To show how expressive in Hindi are novel, one should initially inspect the general examples of reduplication in the language. Mohan (2006, 119-120) portrays the accompanying reduplication designs in Hindi:

1) Full lexical reduplication: constituted by the reduplication of lexemes to present durative or iterative conditions.

baiṭhe-baiṭhe 'while sitting'

da:ne 'grain' *da:ne-da:ne* 'each grain'

2) Partial lexical reduplication (echo-formation): When the initial consonant of the base is replaced by 'v' in the reduplicate, to give a meaning of 'etc.'

kha:na: 'food' *kha:na:-va:na*: 'food, etc.'

roti 'flatbread' *roti-voti* 'flatbread, etc.'

3) Redundant Compounds: Two words placed together that have the same semantic meaning, and some shared phonological qualities, imparting the added quality of 'etc.'

dhān 'wealth' *dhān-dāulət* 'wealth-wealth'

dhārm 'religion' *dhārm-ima:n* 'religion-religion'

4) Expressive reduplication: "minimally meaningful and segmental indivisible morphemes which are constituted of iterated syllables" (Mohan 2015, 119).

ṭəp-ṭəp 'sound of water dripping'

cit-cit 'sticky'

Mohan's examples show only the most basic expressive form, in which a monosyllabic base is copied in the reduplicant. However, expressive in Hindi show a widely variable structure, significantly differing from the regimented types of reduplication in the prosaic grammar (1-3). For example:

1) C₁VC-C₂VC

- *rim-jhim* 'sound of rain falling'
- *jhil-mil* 'shining'

2) CV₁C-CV₂C

- *dhak-dhuk* 'heartbeat'
- *ʈa:pur-ʈupur* 'rain falling'

3) C₁V₁C₂V₁C₂V₁-C₁V₁C₂ (V₁)

- *jhānānā-jhān* 'sound of foot anklets'
- *ghānānā-ghānā* 'sound of thunder'

Expressive in this manner show an immense range of conceivable phonological and morphological examples that withdraw essentially from the fixed examples of full, incomplete and excess reduplication in Hindi as delineated by Mohan. The sheer broadness of structure and the changeable capacity to avoid characterization as for phonological and morphological designing proposes that expressives ought to be considered as a different word class in Hindi, and inspected separated from dull reduplication.

Since its beginning in the mid 20th century, tune has been an inseparable piece of Hindi film. However, as Sarasin notes, film melodies are not just an element of film, they are the major mainstream melodic classification in South Asia. Inside the movies themselves melodies are sung corresponding to numerous subjects, for example, the festival of occasions, for example, marriage or social celebrations or to make individuals excited for some reason, for example, fight. In any case, the most mainstream subject to event a tune inside Hindi film is that of sentimental love.

Expressive serve a specific capacity here, for it is in articulations of affection in which representation, metonymy and other beautiful gadgets are utilized most regularly, and for which the syntactic adaptability of the expressive word class is particularly fit. For example, expressives are ordinarily heard related to explicit items, for example, a lady's anklets (pa:yəl) which through the activities of dance, pass on (or 'talk') feelings that are not suitable to state in words.

Expressives additionally encourage feeling loaded relations between the normal world, for example, climactic highlights, for example, downpour or the shimmering of stars in the night sky, and the darling and their adored. At last, it is basic for specific pieces of the body, for example, the 'heart' or the 'eyes' to metonymically sub for the human individual themselves, imparting an entire scope of feelings, a beautiful element additionally encouraged by the utilization of expressives. However, Hindi film tune verses likewise show an innovative exchange between the two of every a cycle I call here "expressivization." Expressivization happens when notable or phonaesthetic components in mundane action word structures are reworked as expressives and the other way around.

In Hindi, expressivization happens when the verbal base is reduplicated Hindi film melodies contact a huge, semantically different crowd, and the liberal utilization of expressives recommends that expressive implications are generally surely known even to non-local Hindi speakers. As this volume attempts to show, expressives have open notability all through the South Asian etymological zone overall. As of late there has been a propensity for Hindi film melodies to fuse in a more considerable way territorial phonetic assortments.

Since expressives are expected to have cross-semantic conceivability, they frequently show up in such multilingual tunes. For example, the fight tune arrangement in the new move Bajirao Mastani fuses various Marathi structures, however expressives are utilized to limit semantic distinction and forefront the graceful components. The Marathi structures are underlined.

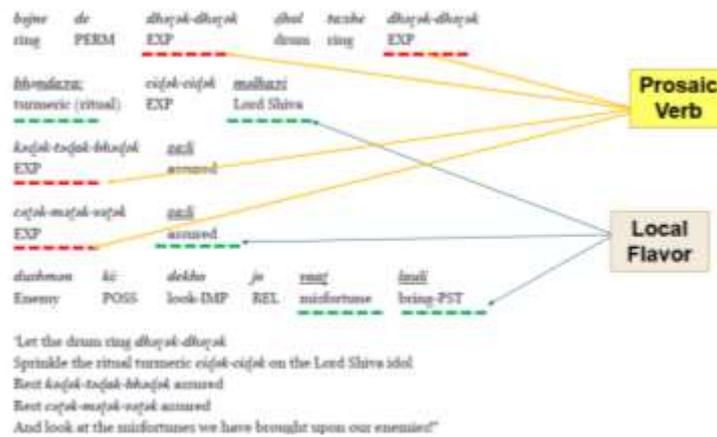


Figure 1. Expressivization of prosaic verb and local flavour in Malhari of movie Bajirao Mastani

5. Conclusion

The liberal utilization of expressives in Hindi film tunes since the initiation of Hindi film in the mid 20th century makes them an ideal scene to examine expressives. The tune verses and groupings in which they happen show the linguistic scope of these structures, just as the words' tasteful impacts and interrelation with dance, music, verse, and realistic visuality. Reduplication in Hindi is an ordinary wonder, introducing designing along the lines of full reduplication (for emphasis), halfway reduplication (reverberation development) or excess compounding. The melody verses and successions in which they happen show the linguistic scope of these structures, just as the words' stylish impacts and interrelation with dance, music, verse, and artistic visuality. From 2006 and up till now, the Hindi tunes has encountered an adjustment in the phonetic highlights, one of the fundamental factors that add to this change is a direct result of globalization.

For a change, lyricists have decided to remember some English words for their tunes since Indians presently comprehend this unknown dialect. One heading for future work is to investigate other melodic societies worldwide just as provincial entertainment worlds in India and stretch out the current dataset to have the option to incorporate more Bollywood tunes. Another heading is that we can chip away at Mauritian Sega to examine the development throughout the long term.

References

- [1] Author A. Frans Wiering, Dept. IT, Utrecht University, Netherlands. "Can Human Benefit From Music Information Retrieval"
- [2] Bigand, E., Poulin-Charronnat, B.: Are We "Experienced Listeners"? A Review of the Musical Capacities That Do Not Depend on Formal Musical Training. *Cognition* 100
- [3] "Simple LRC Format", [http://en.wikipedia.org/wiki/LRC_\(file_format\)](http://en.wikipedia.org/wiki/LRC_(file_format))
- [4] "LRC File Format", [http://en.wikipedia.org/wiki/LRC_\(file_format\)](http://en.wikipedia.org/wiki/LRC_(file_format))
- [5] K. Gajjar and S. Shah. Article: Mood based playlist generation for hindi popular music: A proposed model. *International Journal of Computer Applications*, 2015.
- [6] F. Kleedorfer, P. Knees, and T. Pohle. Oh oh oh whoah! towards automatic topic detection in song lyrics. In *Ninth International Conference on Music Information Retrieval*, 2008, pages 287–292, 2008.
- [7] M. Choudhury, R. Bhagwan, and K. Bali. "The Use of Melodic Scales in Bollywood Music: An Empirical Study," Microsoft Research Lab India, *International Society for Music Information Retrieval*, 2013.
- [8] B. Patra, D. Das, and S. Bandyopadhyay, "Automatic Music Mood Classification of Hindi Songs," [Proceedings of the 3rd Workshop on Sentiment Analysis where AI meets Psychology (SAAIP 2013), IJCNLP 2013, pages 24–28, Nagoya, Japan, 2013
- [9] Harika Abburi, Eswar Sai Akhil Akkireddy, Suryakanth V Gangashetty, Radhika Mamidi Language Technology Research Center IIIT Hyderabad India, *Multimodal Sentiment Analysis of Telugu Songs Proceedings of the 4th Workshop on Sentiment Analysis where AI meets Psychology (SAAIP 2016), IJCAI 2016, pages 48-52, New York City, USA, July 10, 201*

- [10] Amina Abdul Shakoor, Waffiqah Bibi Sahebodin, Sameerchand Pudaruth, Exploring the Evolutionary Change in Bollywood Lyrics over the Last Two Decades ISBN: 978-1-941968-13-0 ©2015 SDIWC Proceedings of the Second International Conference on Data Mining, Internet Computing, and Big Data, Reduit, Mauritius 2015
- [11] XIA, Y., Wang, L., WONG, K.-F., & Xu, M. (2008). Lyric-based Song Sentiment Classification with Sentiment Vector Space Model. *International Journal of Computer Processing of Languages*, 21(4),309330.<https://doi.org/10.1142/S1793840608001950>
- [12] Oudenne, A. M., & Chasins, S. E. (2015). Identifying the Emotional Polarity of Song Lyrics through Natural Language Processing. *Cpsc65*, 1–14.
- [13] Corona, H. (2015). An Exploration of Mood Classification in the Million Songs Dataset, (August).
- [14] Shanmugapriya, K., & Srinivasan, B. (2015). An Efficient Method for Determining Sentiment from Song Lyrics Based On WordNet Representation Using HMM. *Ijirce.Com*, 1139–1145. Retrieved from http://www.ijirce.com/upload/2015/february/96_AN.pdf
- [15] Ascalon, E. I. V, & Cabredo, R. (2015). Lyric-Based Music Mood Recognition, 3, 1–8.
- [16] Sharma, G., & Murty, M. N. (2011). Mining Sentiments from Songs Using, 328–339.
- [17] Walsten, D., & Orth, D. (n.d.). Song Genre Classification through Quantitative Analysis of Lyrics.
- [18] Vijayakumar, C., & Adaickalam, V. (2016). Smart Stress Relieving Music Player using Intelligent Sentiment Analysis 20–23.
- [19] Atia Anwar Zoon*, Dr. Ghulam Ali Buriro†, Muhammad Abdullah‡, Dr. Aazadi Fateh Muhammad§, GENDER ROLES AND REPRESENTATIONS IN BOLLYWOOD ITEM SONGS: A CRITICAL DISCOURSE ANALYSIS OF EMERGING TERMS The Women, *Research Journal*, Volume 11, 2019 Page 139-155
- [20] Choksi, N. (2020). Expressives in Hindi Language Film Songs. In *Expressives in the South Asian Linguistic Area*, Nathan Badenoch and Nishaant Choksi, eds. (pp. 177-194). Leiden: Brill. Author Accepted Manuscript)
- [21] Mohan, S. (2006). Echo-Word Formation in Hindi. *Indian Linguistics* 67: 119-126.
- [22] Abbi, A. (1985). Reduplicative structures: A phenomenon of the South Asian linguistic area. *Oceanic Linguistics Special Publications* 20:159-171.
- [23] Emmeneau, M. (1980). *Language and linguistic area*. Stanford: Stanford University Press.
- [24] Jakobson, R. (1960). Linguistics and poetics. In *Style in language*. (pp. 350-377). Cambridge MA: MIT Press