Turkish Online Journal of Qualitative Inquiry (TOJQI)
Volume 12, Issue 6, July, 2021: 8866 - 8873

Knowledge Discovery of Tones In Songs Using Bi-Lstm

Dr. Pankaj Lathar

Dept. Of Computer Science & Applications
Bhai Parmanand Institute Of Business Studies,New Delhi,India
(Now Dseu Shakarpur Campus)

Abstract

Music Analysis And Songs Recommendation Is An Interesting Area Of Computer Science Research. The Knowledge Discovery Of Songs Involves The Analysis Of Various Tones In The Songs. The Tones Can Be Joy, Sad, Confident And Others. The Analysis Of The Tones In The Songs Can Enable The User To Customize The Playlist Based On The Mood Of The User. Thus, It Is Essential For Enhancing The Techniques For Knowledge Discovery Of Tones In Music Analysis. In This Paper, Music Cognitive Agent (Mca) Is Proposed For The Analysis Of Tones In Songs. The Cognitive Environment Consists Of Song Discovery, Playlist Analysis And The Tone Analysis As A Part Of The Knowledge Discovery. The Proposed Mca Is Compared With Neural Network Bi-Lstm To Compare The Discovery Of The Tones In The Cognitive Environment. The Results Suggest That The Mca Recognizes The Tones Appropriately.

Keywords: Music Analysis, Cognitive Analysis, Playlist Analysis, Tone Analysis.

Introduction

In Recent Years, Production And Consumption Of Music Has Increased Manifold Owing Largely To Effortless Access. Social Messaging Applications Such As Whatsapp, Line, Wechat, Video Sharing Sites Such As Youtube, Vimeo, Prime, Metacafe And Social Networking Sites Such As Facebook, Twitter, Have Billions Of Monthly Active Users Who Share, Interact With And Consume Various Types Of Content, Including Music. Apps Such As Spotify, Youtube Music, Wynk, Apple Music Etc., Focus Primarily On Personalising This Listening Experience For Their Users. These Applications Try To Derive Music Preferences Based On Listening Patterns And Use It As An Important User-Retention Tool. While Language And Genre Preferences Are Common Ways To Tailor Playlists For Listeners, It Is A Well-Known Fact That Choice Of Music Also Depends Heavily On The Listener’s Mood. Thus, The Ability To Classify Songs Into Different Emotional Categories Is An Important Step In Furthering Personalization And Enhancing User Experience. Lyrics Plays An Integral Part To Understand The Tone Of The Music For A Peculiar Song. One Of The Most Challenging Task Is To Find Touching Features In Song Lyrics To Label It As Emotional States. When People Listen To Cheerful Music, Chemicals Like Dopamine And Serotonin Are Generally Produced By The Brain. Likewise, Different Emotions Are Caused When People Listen To Soothing, Hard, Loud Or Angry Music. While Listening To Music, A Person May Feel Ecstatic, Sad, Annoyed Or Relaxed. Music May Cause An Individual To Feel Or A Combination Of These Emotions. (Al Marouf Et Al., 2019)

Music Is Ubiquitous That Can Clout Human Attention (Herget, A. K. Et Al., 2018). Bustling Guitar Sounds Or Piano Themes Might Influence Our Behavior And Emotions. In Past Few Decades, Several Studies Showed That When Combining The Images And Music In Audiovisual Media Can Affect The Cognition And Behavior. Integrating Music Into Visual Media Content As A Way To A Specific End. Nowadays, Musical Accompaniment Is Used By Advertisers To Augment Their Promotional Messages. Producers Of Tv Shows And Movies Often Applies Music As An Intensifier Of The Anecdote Course. Music Is Considered As A “Catalyst Of Advertising”, As It Provides An Integral Support For Advertisements In Commercial Breaks. Pertinent Music “Augments Pictures And Colors Words, And Usually Adds A Form Of Energy Available

Through No Other Source” Conclude That People Nowadays Are Getting More Involved In Advertising, So Music Seems To Be A Perfect Vehicle To Convey The Advertising Messages.

In This Paper, A Cognitive Based Methodology Has Been Proposed To Analyze The Tones Of A Specific Song Lyrics And A System To Label Song Tones Into Appropriate Classes. The Rest Of The Paper Is Organized As Follows. In Section Ii The Related Works In Tone Analysis Are Discussed And Followed By The Proposed Methodology In Section Iii. In Section Iv The Experimental Results Found By The Proposed System Is Discussed And Conclusion In Section V.

Related Work

The Application Aims At Creating A Solution For Tagging Music Based On Genre, Theme Of The Song, And Emotions In The Lyrics. There Are A Variety Of Different Approaches That This Task May Be Carried Out And We Aim To Find The Most Efficient And Effective Implementation. Music Playlist Creating Models Are Already Implemented And Are Used On Various Applications Like Spotify And Apple Music. These Models Tag The Music Based On Genre And Create Playlists Of Songs Based On The Taste Of The User’s Listening Choice. This Model Is Similar To Google Ad-Sense And Netflix’s Recommendation Engine. The Software Aims Are Enhancing The Tags Available For A Particular Piece Of English Music By Extracting The Themes And Sentiment Associated With It. The Main Stakeholders In This Engine Would Be The End User, Who Can View The Theme And Mood Of A Song Quickly, As Well As Other Apis Which Can Use The Extra Tags We Generate In Order To Create Better Playlist’s Suited To The User’s Listening Habits. The Software Is Divided Into Parts And Each Part Implements A Feature. The Scope Of This Project Is To Implement The Features Necessary To Identify The Title Of The Song, Identify The Genre Of Music, Grab Sentiment From The Lyrics, And Find The Main Themes Being Depicted.

An Acoustic Fingerprint Is A Fingerprint Deterministically Generated From An Audio Signal. It Is A Condensed Digital Summary That Can Be Used To Determine An Audio Sample. An Acoustic Fingerprint Quickly Locate Analogous Items In An Audio Database. In Our Implementation We Must Be Able To Retrieve The Title Of The Input Track Quickly With A High Level Of Accuracy In Order To Obtain The Lyrics For The Song To Carry Out Further Analysis. Audio Fingerprinting Is A Process By Which Music Is Identified With A Small Sample Of The Track To Be Identified. Audio Is Identified Independently Of Its Format By The Technologies Like Audio Fingerprinting Without The Need Of Meta-Data Or Watermark Embedding. There Are Many Other Uses Of Fingerprinting Like Watermark Support, Integrity Verification And Content-Based Audio Retrieval. Fingerprinting Has Several Approaches That Have Been Described With Different Rationales And Terminology: Pattern Matching, Multimedia (Music) Information Retrieval Or Cryptography (Robust Hashing) (Cano, P. Et Al, 2005)

The Application Shall Go About The Implementation Of Song Fingerprinting By Using The Echoprint Service. (Echoprint) Is A Music Identification System Build By The (Echo Nest), Which Utilizes Acoustic Fingerprinting Technology For The Identification Functionality. Think Of Actual Human Fingerprints Being Used For Identification. Acoustic Fingerprinting Uses The Same Principle, But By Means Of Audio. Echoprint Consists Of A Set Of Components Which Can Be Used To Either Experiment With And/Or Set-Up An Audio Identification System/Service. An Acoustic Fingerprint Is Achieved By Creating A Condensed Digital Summary Of The Audio Signal. The System “Listens” To The Audio Signal Being Played. An Algorithm Places Marks On, For Example, Different Spikes In Frequency Of The Signal And Is Able To Identify The Signal By Matching These Marks To A Database On An External Server. The Signal Can Include All Different Forms Of Audio, Songs, Melodies, Tunes And For Example Sound Effects From Advertisements. The Ability Of Echoprint’s To Identify A Music Track Is Substantially Fast With Great Efficiency That Makes It One Of The Most Valuable Music Identification Systems Available. Additionally, It Can Even Recognize Noisy Versions Of The Original Track And Even Recordings Performed By Mobile Devices With Noise “Bleed” By Environmental Factors (Brinkman, C. Et Al., (2016)

The Next Phase Involves The Identification Of The Genre Of Music. By Passing The Song Through A Machine Learning Model Can Analyze The Genre Of The Song To A Very High Veracity. The Optimal Class Boundaries Are Attained By The Support Vector Machines Between Different Genres Of Music By Learning From Training Data. Experimental Results Of Multi-Layer Support Vector Machines Illustrate Good Performance In Musical Genre Classification And Are More Advantageous Than Traditional Euclidean

Distance Based Method And Other Statistic Learning Methods (Xu, C. Et Al., 2003). The Implementation Discusses Relevant Features To Identify The Genre Like Beat Spectrum, Lpc (Linear Predictive Coding, Zero Crossing Rates, Spectrum Power And Mel Frequency Cepstral Coefficients.)

Recognizing The Emotional States Like Sorrow, Happiness, Joy Etc Of A Person Is An Uphill Task. To Learn Emotions From Textual Data, (Alm Et Al., 2005) Suggested Arrangement Of Sentences From Fairy Tales. Machine Learning Based Models Applied For The Prediction Task Is Presented In (Alm Et Al., 2005). 14 Defined Features Are Extracted From These Stories And A 10-Fold Cross Validation Is Made To Assess The Accuracy.

Texts In Literary Works Having Different Flavors And Artistic Touches Are Fed As Inputs To Extract The Emotional States. Read Has Classified Short Stories Which Have Sentence-Level Emotion Annotations, On The Other Hand Common-Sense Knowledge Base Is Used In (Liu, H. Et Al., 2003) To Depict Real-Life Knowledge About Emotions. However, Textual Based Identification Of Emotions Is A Challenging Task And The Methods Applied Have Their Own Limitations. Some Of Them Prescribed In (Wu, C. H. Et Al., 2006) Are: "Ambiguity In Keyword Definitions, Incapability Of Recognizing Sentences Without Keywords, Lack Of Linguistic Information, Difficulties In Determining Emotion Indicators Etc." (Al Marouf Et Al., 2019)

The Project Intends To Approach The Sentiment Analysis In Two Methods, One Would Be Using The Watson Api For Tone Analysis (Gundecha, P. (2016). This Returns 6 Tones In Speech As Well As The Degree To Which The Tone Is Positive Or Negative. First, A Method Is Proposed For Target Representation That Captures The Semantic Meaning Of The Opinion Target. Second, An Attention Model Has Been Proposed That Incorporates Syntactic Information Into The Attention Mechanism (He, R., Lee Et Al., 2018).

The Last And Final Phase Of The Project Includes The Topic Extraction. As It Is Difficult To Understand The Type Of Music We May Come Across, It Is Imperative That We Have An Unsupervised Model, Which Will Be Able To Pick Out The Main Topics In A Song Regardless Of The Content. Soft Clustering Is Used By An Unsupervised Topic Model Over Distributed Representations Of Words. By Using A Log-Linear Model, The Distributed Word Representations Are Obtained. We Model The Low-Dimensional Semantic Vector Space Represented By The Dense Word Vectors Using Gaussian Mixture Models (Gmms). Even Though The Power Of Lda Algorithm Has Been Extensively Used For Leveraging Topics, Very Few Studies Have Been Reported For Mapping These Statistically Outputted Topics To Semantically Rich Concepts. The Proposed Framework Is An Attempt To Address This Issue By Making Use Of Lda Algorithm To Develop Topics. By Using A New Statistical Weighting Scheme, Project Leverages Concepts From Such Topics And Some Lightweight Linguistic Processes (Anoop, V. S. Et Al., 2016)

By Integrating The Four Above Features The Project Aims At Creating An Engine That Will Help Music Listeners Worldwide Quickly Make Decisions On Whether To Listen To A Song Or Not Based On It's Content And Genre. The Engine Can Also Benefit Already Existing Models Which Try To Curate Playlists.

Proposed Approach

The Proposed Approach Mca Consists Of Four Modules Namely Playlist Analysis Environment, Song Discovery Environment And Tone Analysis Environment. The Overall Proposed Approach Is As Shown In The Figure 1.

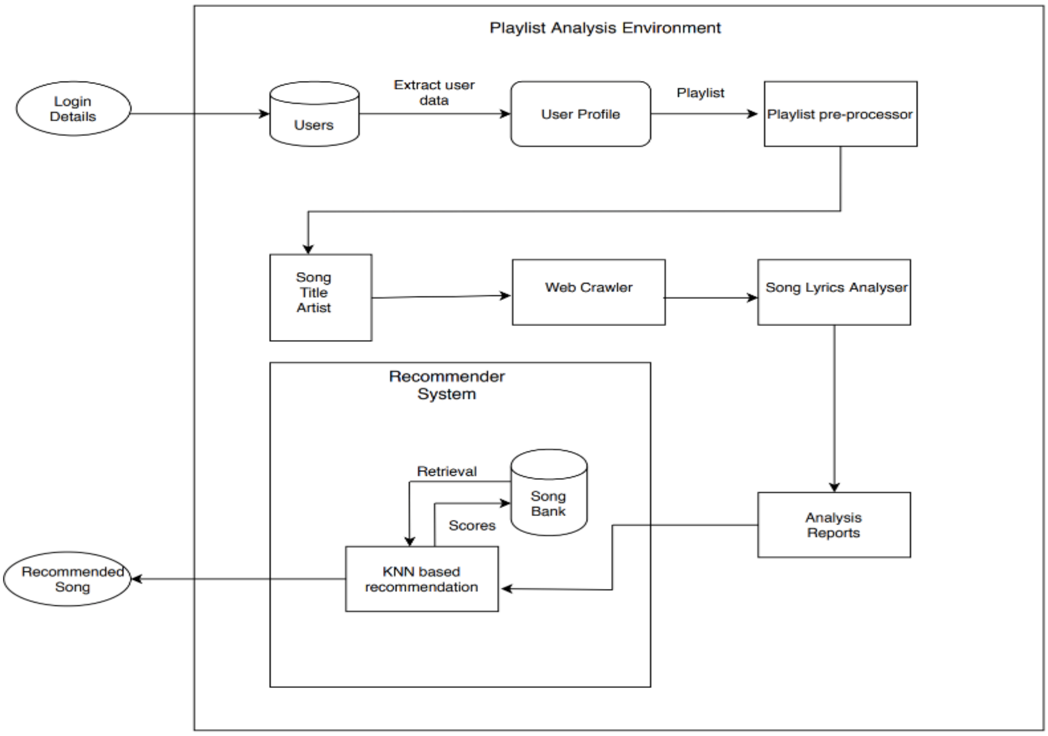

Figure 1. Proposed Mca System

The Sequence Of The Operations Performed By The Mca Is As Shown In The Figure 2. The Audio File Is First Parsed Using The Audio Content Recognition (Acr) For Extracting The Name, Title, Artist Of The Songs. These Data Are Used As A Metadata For The Tone Analysis. The Lyrics Of The Song Are Extracted By The Web Crawler Using The Name And Artist Data Captured Using Acr. The Song Lyric Analyzer Function Is To Determine The Tone Analysis. The Analysis Of The Tones Is Carried Out Using The Watson Tone Analyzer, Latent Dirichlet Allocation (Lda) And Vader Polarity Analyzer. The Details Of The Extraction Of The Tone Analysis Is Discussed In Detail In The Further Sections.

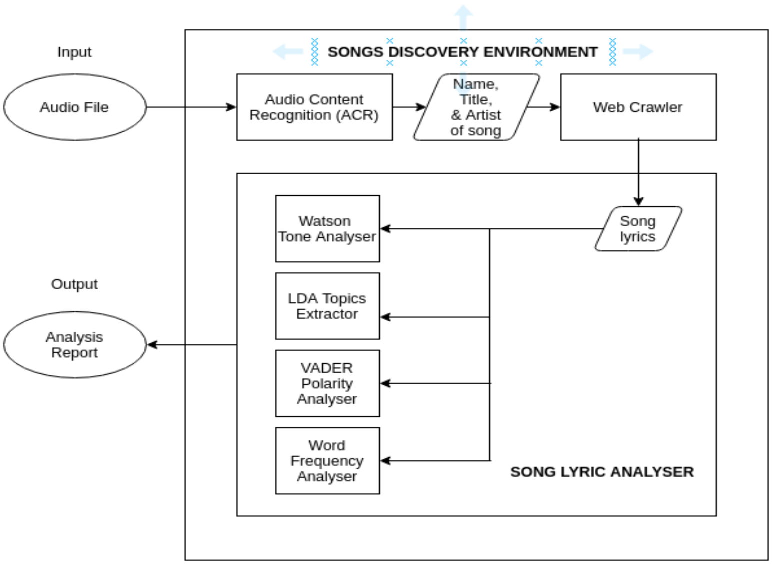

Figure 2. Operations In Mca For Tone Analysis

Playlist Analysis Environment

Allows The User To Login To A Personal Space Wherein The User Can Store Songs In A Playlist Which Works With A Firebase Backend Database. The Tone Scores For The Playlist Songs Are Obtained And Analysed By Comparing With The Top 100 Songs For Each Of The Years In The Past Two Decades. Upon Comparison Of The Euclidean Distance Of The Score Vectors Of The Playlist Songs And Normalized Scores Of Each Of The Songs From The 2000 Songs Data, We Can Find The Nearest Neighbor Vectors To The Playlist Songs And Recommend Similar Scored Songs From Top Songs Data. Hence, Implementing A Recommendation System Based On Emotional Content Of The Lyrics Of The Song Rather Than The Music Itself. The Playlist Can Be Manipulated As Per The Users Will And May Or May Not Add Recommended Songs To It.

Songs Discovery Environment

First Records A Song From The User. The Recorded File Is Sent To The Acr Cloud Recognizer Which Provides The User With The Song Title And Other Details Of The Song. These Details Are Fed To A Web Crawler Implemented Using Beautiful Soup Libraries. The Web Crawler Gives Us The Lyrics Of The Song Under Focus. The Lyrics Once Obtained Are Used In 4 Different Ways:

1. Wordcloud Of The Song Lyrics Can Be Observed, Constructed Using Certain Natural Language Tools Of Python Libraries.
2. The Dominant Tones Reflected By The Lyrics Can Be Observed By Passing Them Through The Tone Analyser Api.
3. The Dominant Topics Underlying In The Lyrics Are Brought Out Using A Latent Dirchlet Allocation Model Trained On 250,000 Songs Labelled With Their Topics.
4. A Polarity Graph Using The Vader Nltk Model Was Constructed To See How Positive Or Negative The Song Was Oriented.

Lda Is The Most Commonly Used Topic Modeling Method As Shown In The Algorithm 1. It Is An Unsupervised Generative Probabilistic Method For Modeling A Corpus. Lda Assumes That Each Document Can Be Represented As A Probabilistic Distribution Over Latent Topics, And That Topic Distribution In All Documents Share A Common Dirichlet Prior. In The Lda Model, Each Latent Topic Is Represented As A Probabilistic Distribution Over Words And The Word Distributions Of Topics Share A Common Dirichlet Prior As Well. Given A Corpus D Consisting Of M Documents, With Document D Having N D Words (D ϵ {1,..., M}), Lda Models D According To The Following Generative Process.

Algorithm 1

Lda Algorithm

(A)Choose A Multinomial Distribution Φ_t For Topic T (T $\in$ {1,..., T}) From A Dirichlet Distribution With Parameter B.

(B) Choose A Multinomial Distribution Θ_d For Document D (D $\in$ {1,..., M}) From A Dirichlet Distribution With Parameter A.

(C)For A Word W_n (N $\in$ {1,..., N_d }) In Document D,

    (I) Select A Topic Z_n From Θ_d.

    (Ii) Select A Word W_n From Φ Z_n.

In Above Generative Process, Words In Documents Are The Only Observed Variables While Others Are Latent Variables (Φ And Θ) And Hyper Parameters (Α And Β). In Order To Infer The Latent Variables And Hyper Parameters, The Probability Of Observed Data D Is Computed And Maximized As Shown In The Equation 1:

$$p(D|\alpha,\beta) = \prod_{d=1}^{M} \int p(\theta_d|\alpha) \left(\sum_{n=1}^{N_d} p(z_{dn}|\theta_d) p(w_{d_n}|z_{d_n},\varphi) P(\varphi|\beta)\right) d\theta_d d_p \quad (1)$$

Tone And Sentiment Analysis

This Module Provided Window Where A Series Of Analytical Reports And Collected Data Regarding Songs Passed Through The Tone Analyser Are Presented In An Interactive Graphical Interface. This Data Shows The Scores Of The Different Sentiments Or Tones Of The Song Lyrics, Collected Over 20 Years. This Was Done By Formatting A List Of The Songs And Their Artist Names Into A Text File And Passing That Through The Tone Analyser Watson Api. Vader Is Used For The Sentimental Analysis Of The Tones.

Vader Is A Lexicon With Both Polarity And Intensity Information Attached To Each Entry. The Lexicon Only Contains Entries With A Standard Deviation Less Than 2.5. Based On The Information We Observed, The Polarity Of Each Word Entry W Can Be Determined As Shown In The Equation 2. Vader Provides Unigram Word Entries As Feature. To Be Used For Feature Scoring Algorithm, The Scoring Function Is Designed As Shown In The Equation 3. This Way Vader Is Prepared To Be Used In Lexicon-Based Sentiment Classification Tasks.

$$pol_{vader}(w) = \begin{cases} 1, \text{if } Intensity(w) > 0, \\ -1, \text{if } Intensity(w) < 0. \end{cases} \quad (2)$$

$$score_{vader0}(w, l) = Intensity(w). \quad (3)$$

Experiments And Results

In This Section The Experiments And Results Are Discussed. The Dataset Consists Of 2000 Songs That Are Over The Years From 2000-2019 Collected Over The 2 Decade Containing The Lyrics Of Each Songs. It Was Acquired Through World Wide Web (Www) Using The Keywords And Meta Searching. For Example, ‘English Sad Songs’, ‘English Happy Songs’ Was Considered As The Keywords For Collecting The Songs. Table 1 Shows Some Of The Properties Of The Collected Dataset.

Table 1. Properties Of Dataset

Properties	**Songs**
No Of Songs	2000
Avg Words Per Song	223.89
Total Words	4,42,313.33

Dataset

Input Data Consists Of Song Lyrics As Well As The 8 Most Common Tones; Analytical, Anger, Confidence, Fear, Joy, None. Sadness, Tentative. The Softmax Output Of The Model Was Used To Generate The Scores Used For Comparison. In The Application We Display The Top 3 Labels, However For The Purpose Of Comparison We Will Be Concentrating Only On The Label With The Highest Probability.

Results And Comparison

The Class Scores For Each Of The 2000 Songs Was Measured And Used For The Classification Of The Tones. The Results Of The Statistical Measurements Are As Shown In The Table 2.

Table 2. Statistical Measurements Of Tone Class Scores

Tone Class	**Class Percentage In Dataset**	**Minimum Score**	**Maximum Score**
Analytical	19%	0.5	0.9837
Confident	34%	0.5	0.942
Tentative	47%	0.5	0.865

We Built A Deep Learning Model Using Pytorch In Order To Compare The Output Of Ibm Watson's Tone Analyser. The Model Consisted Of A Two Layer Bilstm Followed By 2 Fully Connected Layers With A Hidden Size Of 256. The Learning Rate Used Was 0.001 Along With Cross Entropy Loss Whose Weights Were Initialized Inversely Proportional To The Class Imbalance Present In The Dataset. The Model Was Trained For 5 Epochs With Around 2000 Data Points. The Comparison Of The Analysis With Bilstm And Ibm Watson Model Is As Shown In The Table 3.

Table 3. Comparison Between Bilstm And Ibm Watson Model

Song Title	**Ibm Watson Label**	**Probability Of Label (Ibm Watson)**	**Bilstm Model Label**	**Probability Of Label (Bilstm)**
Beyone's I'm Alone Now	**Tentative**	**0.892**	Confident	0.672
Closure's Look Out Below	**Confident**	**0.663**	Confident	0.543
Forever Tonight's Holding On	Sadness	0.6539	Sadness	.8937
Funkadelic's Balance	None	1	Joy	.6432
Diamond Rio's Start All Over Again	**Confident**	**0.875**	Confident	0.592

Upon Further Analysis Of The Bilstm Model We Came To The Following Conclusions:

- The Model Had A Hard Time Selecting Out The 'None' Ie, No Clear Tone And Ended Up Defaulting To Classes Like Joy And Sadness Which Had Higher Samples In The Training Data.
- The Model Also Had Issues Selecting Out The Difference Between Tentative, Confident, And Angry.
- We Believe The Reliability Of The Bilstm Model Can Be Improved With A More Complex Model And More Training Data.

Conclusion

The Overall Implementation Of The Project Has Been Made Possible By Developing Each Module Individually. The Algorithm Which Describes The Flow And The Working Of The Project Shows How The User Can Navigate Through The Application And Find Various Useful And Interesting Results Done From

Studies And Analysis. The Different Models Are Used To Achieve These Results. The Reasons For Using Nltk Vader For Polarity Charts Include Accuracy Of The Model And Efficient Working. The Utilization Of Lda Can Be Justified Due The Speed At Which This Model Provides Results. The Nearest Neighbour Vector Recommendation System Based On Euclidean Distance Is The Best Method To Find Similar Songs One May Like Using Space Vectors. From The Above Analysis We Can Conclude That We Have Started A Small Application Which Can Prove To Be Beneficial To Music Lovers And Composers In The Future. The Application Is Able To Identify A Piece Of Music, Extract The Tones, Topics And Give Some Visualization On The Same. This Helps Give A Quantification To Textual Data Embedded Within The Lyrics Of A Song. These Data Points Can Be Used In The Future To Carry Out Further Analysis.

References

1. Al Marouf, A., Hossain, R., Sarker, M. R. K. R., Pandey, B., & Siddiquee, S. M. T. (2019, February). Recognizing Language And Emotional Tone From Music Lyrics Using Ibm Watson Tone Analyzer. In 2019 Ieee International Conference On Electrical, Computer And Communication Technologies (Icecct) (Pp. 1-6). Ieee.
2. Herget, A. K., Schramm, H., & Breves, P. (2018). Development And Testing Of An Instrument To Determine Musical Fit In Audio–Visual Advertising. Musicae Scientiae, 22(3), 362-376.
3. Cano, P., Batlle, E., Kalker, T., & Haitsma, J. (2005). A Review Of Audio Fingerprinting. Journal Of Vlsi Signal Processing Systems For Signal, Image And Video Technology, 41(3), 271-284.
4. Echoprint Https://Echoprint.Tumblr.Com/
5. Echonest Https://Github.Com/Echonest
6. Brinkman, C., Fragkiadakis, M., & Bos, X. (2016). Online Music Recognition: The Echoprint System.
7. Xu, C., Maddage, N. C., Shao, X., Cao, F., & Tian, Q. (2003, April). Musical Genre Classification Using Support Vector Machines. In 2003 Ieee International Conference On Acoustics, Speech, And Signal Processing, 2003. Proceedings.(Icassp'03). (Vol. 5, Pp. V-429). Ieee.
8. Alm, C. O., Roth, D., & Sproat, R. (2005, October). Emotions From Text: Machine Learning For Text-Based Emotion Prediction. In Proceedings Of The Conference On Human Language Technology And Empirical Methods In Natural Language Processing (Pp. 579-586). Association For Computational Linguistics.
9. Liu, H., Lieberman, H., & Selker, T. (2003, January). A Model Of Textual Affect Sensing Using Real-World Knowledge. In Proceedings Of The 8th International Conference On Intelligent User Interfaces (Pp. 125-132).
10. Wu, C. H., Chuang, Z. J., & Lin, Y. C. (2006). Emotion Recognition From Text Using Semantic Labels And Separable Mixture Models. Acm Transactions On Asian Language Information Processing (Talip), 5(2), 165-183.
11. Gundecha, P. (2016). Ibm Watson Just Got More Accurate At Detecting Emotions. Https://Www. Ibm. Com/Blogs/Bluemix/2016/10/Watson-Has-More-Accurate-Emotion-Detection/. Último Acesso, 19(04), 2019.
12. He, R., Lee, W. S., Ng, H. T., & Dahlmeier, D. (2018, August). Effective Attention Modeling For Aspect-Level Sentiment Classification. In Proceedings Of The 27th International Conference On Computational Linguistics (Pp. 1121-1131).
13. Anoop, V. S., Asharaf, S., & Deepak, P. (2016). Unsupervised Concept Hierarchy Learning: A Topic Modeling Guided Approach. Procedia Computer Science, 89, 386-394.