

Research Article

Sentiment analysis in songs and advanced recommendation system

Mayur Saudagar Saptal and Vanita Raut

Dept. of Computer Engineering, GH Raisoni Institute of Engineering and Technology, Pune, India

Received 10 Nov 2020, Accepted 10 Dec 2020, Available online 01 Feb 2021, **Special Issue-8 (Feb 2021)**

Abstract

To overcome a oversupply problems and recommending a user specific data combine a collaboration filter and content base recommendation in application of sentiment based song recommendation dividing a work load in high end (sever) and low end (mobiles) devices

Keywords: Data mining ; sentiment analysis; recommendation ; Content based ; collation filter ; lexicon ; KNN; text mining

Introduction

Music industry have a oversupply problem to overcome this need to have perfect recommendation algorithm . As well as it in this paper consider a sentiment analysis approach for finding a perfect match of song based on user input. User inputs are get analyses in two type of recommendationType collaborative filter where similar set of data get collected and content based recommendation where are per user profile data get sorted that so combining the both type will lead to a perfect recommendation. Applied this hybrid approach on song recommendation with sentiment analysis for analyzing sediment of song and user input lexicon algorithm is going to used and for collaboration and contend based recommendation K-nearest neighbor (KNN) algorithm is used .

Literature Survey

In the paper “An Analysis of Music Lyrics by Measuring the Distance of Emotion and Sentiment , Jinhyuck Choi^[1], Jin-Hee Song^[1] , Yanggon Kim^[1] ” authors mention the oversupply problem for over come this they had work on lyrics refining and NLP to match the user choice where they analyzed the lyrics emotion using lexicon algorithm and find recommendation using KNN algorithm.

In the paper “Evaluating Sentiment Similarity of Songs Based on Social Media Data, Jiajun Gao^[2], Hua Yuan^[2], Limin Wang^[2], Yu Qian^[2]” authors work on a Classification of comment and extracting the emotions from them for that they gather review of the song from users and on that they create a Classification table and use if it for recommendation.

A. Specifications

For server side PHP is used and front end android (JAVA) or flutter(Dart) is used and data transmission will held on the JSON API which also capable of transmitting media in the chunk.

KNN algorithm is used for classification and recommendation for that “PHP-ML - Machine Learning library for PHP”(<https://github.com/php-ai/php-ml-examples>) library is used which is under MIT Licence. And to extract emotions of the lyrics lexicon algorithm based “php-sentiment-analyzer” (<https://github.com/davmixcool/php-sentiment-analyzer>) library is used which is also under MIT license.

Proposed Methodology

User will give input in text or in speech format mostly in speech format that speech will get converted into a text and that text send to server in json format for analyzed and emotions are get calculated for it as well as keyword search is happened for that that result return millions of songs that songs will again get filter under collaboration filter this result optimized to hundred and that hundred again sent to mobile and based on the user profile it will be sorted in the 20 to 30 songs.

First step the emotions and keyword are consider and based on that most likely data which other users had listened is get extracted that will reduce a cold start problem as we have all data already prepared that data again sorted based on the user contend or profile which consider the most listened genres , musician , song types , music year , duration , listen count , singer and emotions.

A. Architecture

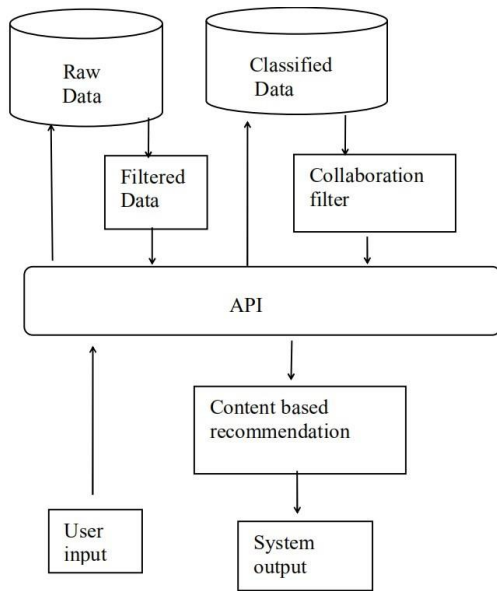


Fig. 1. Architecture diagram

B. Algorithms

KNN algorithm is a supervised machine learning algorithm which is used for classification and regression problem.

Probability:

$$P(Y_i|x) = \frac{N_i + sK}{N + sK + C_s}$$

Where,

X is instant,

N_i is instant belong to Y_i in neighbor,

K is total instant,

C is classes and

S is smooth parameter

Lexicon algorithm is algorithm where all text converted into tokens and tokens have predefined set of scores that score get assigned to the words and using this score complete result or emotion of the text get calculated

C. Proposed

1. Take user input
2. Convert it in text
3. Analyzed it and count emotion score using lexicon algorithm
4. Apply keyword search and emotion based search
5. Apply KNN for collaboration filter on result
6. Send out result to low end device for process
7. On the basic of user profile result get minimize and sorted
8. Shown output to user

Experiment and problems case case 1 : Searching for raw songs data there not many data-sets on internet but i had mange to search some on kagal database and downloaded them in ts , csv and txt files having size

more the than 2gb case 2 : removing special characters from song lyrics

- when analyzed this data found that there and some emoji and song symbols in text so removed them
- case 3 : converting lyrics into utf8 - for purification converted them into a utf8 charters

case 4 : removing unrecognized characters and extra spaces from lyrics

- latter fond that they have some asci character also so need to remove them also as well as remove extra spaces
- case 5 : calculating emotions score of lyrics - this purified text is given to a emotions calculation algorithm

case 6 : storing score , lyrics and meta data of songs in database

-extract the meta , score and purified lyrics for store purpose

case 7 : avoiding maximum connection of mysql database to avoid overloading mysqlserver

- but need to update algorithm again because it is creating to many connection while storing the data

case 8 : taking break after earch 100 songs completion to avoid ram consumption by php - added sleep after every 100 songs

case 9 : avoiding query looping for avoid buffer overflow or server jamming

- created a large query with limited conditions and selection to avoid buffer overflow
- case 10: setting crons for making automation of process - to auto mate it concept of cron is impediment

case 11: setting server of local machine and installing composers

- setup a xampp and installed composer
- case 12: selecting simple framework for use - selected codigniter because it simple and full control code
- case 13: integrating dependencies and library in it - added knn and lexicon lib in project
- case 14: individual method for data filtering based of cron

- raw data filtration in automation process to need to sprearte to it

case 15: converting a file into chunks for filter

- large files processing takes large time and more resource utilization

case 16: file structure and unique id generation to avoid data lose

- to store this chunks need to create a proper dir structure and naming
- case 17: file corruption handling

- avoid name overwriting and currupotion file by adding scanner
- case 18: table structure for data storing

- need to create optimal structure for data stored different types of data in diff rent table
- case 19: writing optimal query to reduce response time

- there are left join in query so optimized them for selected data only
- case 20: ajax implementation

- in web app by using ajax different request and data collection is handled
- case 21: building soft and user-friendly ui - simple list view is better at this stage off

app case 22: creating users user processing part case 23: creating user relation tables
- user rational tables
case 24: getting user inputs in text format
- there are various way user can interact by gesture , by voice or by text but at the end it is converted in text case 25: calculation score of user input and filtering
- this text is used for score calculation but 1st it is filtered for process
case 26: Store every entry and score of text in database
- every users activity is stored for latter analysis
case 27: save song preference of users with meta and score - saved users most visited most listen meta and score case 28: use privies saved data for recommendation - this data is used for recombination with score case 29: limiting the recommendation result
- limited the result to 20 case 30: developing application for android
-Consideration of low end devices case 31: saving user preference in sqlite3 database - no need to save user data on server so it is saved on mobile case 32: making connection to apis - json api is flexible case 33: providing offline user interface - based on user interface it will work offline case 34: optimizing android application response time
- write mobile algo considering the server response data preparation and showing it to users
case 35: selecting technology for android app (flutter,android , hybrid)
case 36: setting development environment for android app development
- installed android studios vs code and setup of paths
case 37: Optimizing json for better api to application interface
case 38: selecting a algorithm for emotions calculation
case 39: base on emotions and preference creating a recombination result
case 40: handling preference

Result and discussions

1 Novelty

Novelty of paper is that creating hybrid approach of combining the content base and calibration filtering base on the one of the selection criteria as key feature in this paper used a sentiment analysis of songs. For content base filter used meta and preferences of user And for calibration filter use a emotion score. By using this approach solved the cold boot problem and improved the quality of recommendation because it consider the user itself and also recommend based on other users with same choice.

2 Future scope

- Graphics driver
For large computation can be use graphic lib than regular system which will improve the machine learning speed in knn
- Speech to text implantation

Current paper consider only text in future by using text to speech we can apply recommendation

1.Gesture

Gesture can be used for emotion calculation

2.Advertising industry

Recommendation is core part of digital advertising 5. Product recommendation

We suggest user related product base on his and other choice currently it is based of collaborative filtering

C. Results

This approach improved the cold boot problem it provides the most matching results to users can be used on cross platforms as user start using the application results get improved.improved recommendation speed and accuracy depend on user choice Some issue are local database always store preference so required size and lower the response speed. And another issue is after the complete preference is set it start recommending same result. But this problem is when database is fixed but in music industry in not be a problem because data is continuous growing .

Conclusions

Hybrid approach of combining the collaboration and content based recommendation definitely improve the recommendation score and most important it solve the cold boot problem as we have similar user already data on server and user base recommendation(content) improve the user interest and made recombination algorithm effective for addition to it we are using users emotions as emotions are directly connected to the songs it will give better user experience to users

As improvement in the mobile devices they have minimum of 1GB of ram and better processor and we are using json for data transfer and java or dart have inbuilt function / library for handle json it make easy to compute hundreds of Result in seconds and showing list. For server php was used it is famous and mostly pre-install on server it don't have library like numpy in paython but it is pretty stable language And large support of library and it is best suitable for lexicon emotions extraction As we are gone use mysql for store the classified data extracted from raw song data it is easy to apply keyword search and emotions based search Latter KNN is used for the classification and collaboration. Future scope of the this approach is we can use it in the advertise recommendation , video recommendation , movies recommendation and product recommendation

References

- [1]. Jinhyuck Choi(Dept. of Computer and Information Sciences Towson University Towson, MD, USA ,jchoi16@students.towson.edu) , Jin-Hee Song (School of IT Convergence Engineering Shinhan University Dongducheon, Gyeonggi-do, Korea jhsong@shinhan.ac.kr) , Yanggon Kim (Dept. of

- Computer and Information Sciences Towson University
Towson, MD, USA
- [2]. ykim@towson.edu)
- [3]. 978-1-5386-5889-5/18/\$31.00 ©2018 IEEE SNPD
2018, June 27-29, 2018, Busan, Korea
- [4]. An Analysis of Music Lyrics by Measuring the Distance of
Emotion and Sentiment
- [5]. Jiajun Gao¹, Hua Yuan^{1,*}, Limin Wang^{2,1}, Yu Qian¹
- [6]. Evaluating Sentiment Similarity of Songs Based on Social
Media Data
- [7]. A Sentiment Analysis Method of Capsule Network Based
on
- [8]. BiLSTM
- [9]. Yongfeng Dong ; Yu Fu ; Liqin Wang ; Yunliang Chen ;
Yao Dong ;
- [10]. Jianxin Li
- [11]. IEEE Access
- [12]. Year: 2020 | Volume: 8 | Journal Article
<https://ieeexplore.ieee.org/document/9007445/>
- [13]. Transfer Correlation Between Textual Content to
Images for
- [14]. Sentiment Analysis
- [15]. Ke Zhang ; Yunwen Zhu ; Wenjun Zhang ; Weilin
Zhang ; Yonghua Zhu
- [16]. IEEE Access
- [17]. Year: 2020 | Volume: 8 | Journal Article
<https://ieeexplore.ieee.org/document/9003301/>
- [18]. MGL-CNN: A Hierarchical Posts Representations
Model for
- [19]. Identifying Depressed Individuals in Online Forums
- [20]. Guozheng Rao ; Yue Zhang ; Li Zhang ; Qing Cong ;
Zhiyong Feng
- [21]. IEEE Access
- [22]. Year: 2020 | Volume: 8 | Journal Article
<https://ieeexplore.ieee.org/document/8998086/>
- [23]. LISA: Language-Independent Method for Aspect-
Based Sentiment
- [24]. Analysis
- [25]. Mohammadreza Shams ; Navid Khoshavi ; Ahmad
Baraani-Dastjerdi
- [26]. IEEE Access
- [27]. Year: 2020 | Volume: 8 | Journal Article
<https://ieeexplore.ieee.org/document/8995576/>
- [28]. Combining goal model with reviews for supporting
the evolution of apps
- [29]. Yuzhou Liu ; Lei Liu ; Huaxiao Liu ; Shanquan Gao
- [30]. IET Software
- [31]. Year: 2020 | Volume: 14, Issue: 1 | Journal Article
<https://ieeexplore.ieee.org/document/8986852/>
- [32]. Reviewer Credibility and Sentiment Analysis Based
User Profile Modelling for Online Product
Recommendation
- [33]. Shigang Hu ; Akshi Kumar ; Fadi Al-Turjman ; Shivam
Gupta ;
- [34]. Simran Seth ; Shubham
- [35]. IEEE Access
- [36]. Year: 2020 | Volume: 8 | Journal Article
<https://ieeexplore.ieee.org/document/8978583/>
- [37]. Aspect-Based Sentiment Classification Using
Interactive Gated
- [38]. Convolutional Network
- [39]. Avinash Kumar ; Vishnu Teja Narapareddy ;
Veerubhotla Aditya Srikanth ; Lalita Bhanu Murthy Neti ;
Aruna Malapati
- [40]. IEEE Access
- [41]. Year: 2020 | Volume: 8 | Journal Article
<https://ieeexplore.ieee.org/document/8972363/>
- [42]. Sentiment Analysis for E-Commerce Product Reviews
in Chinese
- [43]. Based on Sentiment Lexicon and Deep Learning
- [44]. Li Yang ; Ying Li ; Jin Wang ; R. Simon Sherratt
- [45]. IEEE Access
- [46]. Year: 2020 | Volume: 8 | Journal Article
<https://ieeexplore.ieee.org/document/8970492/>
- [47]. Aspect-Level Drug Reviews Sentiment Analysis Based
on
- [48]. Double BiGRU and Knowledge Transfer
- [49]. Yue Han ; Meiling Liu ; Weipeng Jing
- [50]. IEEE Access
- [51]. Year: 2020 | Volume: 8 | Journal Article
<https://ieeexplore.ieee.org/document/8970280/>
- [52]. Multi-Modal Sentiment Classification With
Independent and
- [53]. Interactive Knowledge via Semi-Supervised Learning
- [54]. Dong Zhang ; Shoushan Li ; Qiaoming Zhu ; Guodong
Zhou
- [55]. IEEE Access
- [56]. Year: 2020 | Volume: 8 | Journal Article
<https://ieeexplore.ieee.org/document/8968422/>
- [57]. Sentiment Analysis of Noisy Malay Text: State of Art,
Challenges and Future Work
- [58]. Challenges and Future Work
- [59]. Muhammad Fakhur Razi Abu Bakar ; Norisma Idris ;
Liyana Shuib ;
- [60]. Norazlina Khamis
- [61]. IEEE Access
- [62]. Year: 2020 | Volume: 8 | Journal Article
<https://ieeexplore.ieee.org/document/8967092/>
- [63]. Evaluating Machine Learning Techniques for
Detecting
- [64]. Offensive and Hate Speech in South African Tweets
- [65]. Oluwafemi Oriola ; Eduan Kotzé
- [66]. IEEE Access
- [67]. Year: 2020 | Volume: 8 | Journal Article
<https://ieeexplore.ieee.org/document/8963960/>
- [68]. Variable Convolution and Pooling Convolutional
Neural Network for Text Sentiment Classification
- [69]. Min Dong ; Yongfa Li ; Xue Tang ; Jingyun Xu ; Sheng
Bi ; Yi Cai
- [70]. IEEE Access
- [71]. Year: 2020 | Volume: 8 | Journal Article
<https://ieeexplore.ieee.org/document/8960358/>
- [72]. Context-aware social media user sentiment analysis
- [73]. Bo Liu ; Shijiao Tang ; Xiangguo Sun ; Qiaoyun Chen ;
Jiuxin Cao ;
- [74]. Junzhou Luo ; Shanshan Zhao
- [75]. Tsinghua Science and Technology
- [76]. Year: 2020 | Volume: 25, Issue: 4 | Journal Article
<https://ieeexplore.ieee.org/document/8954871/>
- [77]. Startup Initiative Response Analysis (SIRA)
Framework for
- [78]. Analyzing Startup Initiatives on Twitter
- [79]. Bashayer Alotaibi ; Rabeeh Ayaz Abbasi ; Muhammad
Ahtisham Aslam ; Kawther Saeedi ; Dimah Alahmadi
- [80]. IEEE Access
- [81]. Year: 2020 | Volume: 8 | Journal Article
<https://ieeexplore.ieee.org/document/8954650/>
- [82]. Improving the Performance of Sentiment Analysis of
Tweets Containing Fuzzy Sentiment Using the Feature
Ensemble Model Huyen Trang Phan ; Van Cuong Tran ;
Ngoc Thanh Nguyen ; Dosam
- [83]. Hwang
- [84]. IEEE Access
- [85]. Year: 2020 | Volume: 8 | Journal Article
<https://ieeexplore.ieee.org/document/8949358/>

- [86]. Predicting Supervise Machine Learning Performances for
- [87]. Sentiment Analysis Using Contextual-Based Approaches
- [88]. Azwa Abdul Aziz ; Andrew Starkey
- [89]. IEEE Access
- [90]. Year: 2020 | Volume: 8 | Journal Article <https://ieeexplore.ieee.org/document/8930507/>
- [91]. Feature representations using the reflected rectified linear unit
- [92]. (RReLU) activation
- [93]. Chaity Banerjee ; Tathagata Mukherjee ; Eduardo Pasiliao
- [94]. Big Data Mining and Analytics
- [95]. Year: 2020 | Volume: 3, Issue: 2 | Journal Article <https://ieeexplore.ieee.org/document/9007872/>
- [96]. Text-based price recommendation system for online rental houses Lujia Shen ; Qianjun Liu ; Gong Chen ; Shouling Ji
- [97]. Big Data Mining and Analytics
- [98]. Year: 2020 | Volume: 3, Issue: 2 | Journal Article <https://ieeexplore.ieee.org/document/9007875/>
- [99]. A survey of data partitioning and sampling methods to support big data analysis
- [100]. Mohammad Sultan Mahmud ; Joshua Zhexue Huang ; Salman Salloum ; Tamer Z. Emara ; Kuanishbay Sadatdiyrov
- [101]. Big Data Mining and Analytics
- [102]. Year: 2020 | Volume: 3, Issue: 2 | Journal Article <https://ieeexplore.ieee.org/document/9007871/>
- [103]. Comparative study of statistical features to detect the target event during disaster Madichetty Sreenivasulu ; M. Sridevi Big Data Mining and Analytics
- [104]. Year: 2020 | Volume: 3, Issue: 2 | Journal Article <https://ieeexplore.ieee.org/document/9007873/>
- [105]. A Sentiment Analysis Method of Capsule Network Based on BiLSTM
- [106]. Yongfeng Dong ; Yu Fu ; Liqin Wang ; Yunliang Chen ; Yao Dong ; Jianxin Li IEEE Access
- [107]. Year: 2020 | Volume: 8 | Journal Article <https://ieeexplore.ieee.org/document/9007445/>
- [108]. Personalized Local Event Search Based on SNS Data Analysis Miki Imai ; Miki Enoki ; Ruriko Kudo ; Masato Oguchi
- [109]. 2020 14th International Conference on Ubiquitous Information Management and Communication (IMCOM) Year: 2020 | Conference Paper <https://ieeexplore.ieee.org/document/9001723/>
- [110]. [26] Tactics-Trend Analysis for Increasing the Possibility of Shooting in a Basketball Match
- [111]. Shu Yano ; Kenji Matsuura ; Hiroki Tanioka ; Stephen Karungaru ; Naoya Kohda ; Naka Gotoda ; Tomohito Wada
- [112]. 2020 14th International Conference on Ubiquitous Information Management and Communication (IMCOM) Year: 2020 | Conference Paper <https://ieeexplore.ieee.org/document/9001784/>Data mining with big data
- [113]. Xindong Wu ; Xingquan Zhu ; Gong-Qing Wu ; Wei Ding IEEE Transactions on Knowledge and Data Engineering
- [114]. Year: 2014 | Volume: 26, Issue: 1 | Journal Article Cited by: Papers (1044) <https://ieeexplore.ieee.org/document/6547630/>
- [115]. Big IoT Data Analytics: Architecture, Opportunities, and Open
- [116]. Research Challenges
- [117]. Mohsen Marjani ; Fariza Nasaruddin ; Abdullah Gani ; Ahmad Karim ;
- [118]. Ibrahim Abaker Targio Hashem ; Aisha Siddiqi ; Ibrar Yaqoob
- [119]. IEEE Access
- [120]. Year: 2017 | Volume: 5 | Journal Article <https://ieeexplore.ieee.org/document/7888916/>
- [121]. A Survey of Data Mining and Machine Learning Methods for
- [122]. Cyber Security Intrusion Detection Anna L. Buczak ; Erhan Guven
- [123]. IEEE Communications Surveys & Tutorials Year: 2016 | Volume: 18, Issue: 2 | Journal Article <https://ieeexplore.ieee.org/document/7307098/>
- [124]. Content-Based Superpixel Matching Using Spatially Constrained
- [125]. Student's-t Mixture Model and Scale-Invariant Key-Superpixels
- [126]. Pengyu Wang ; Hongqing Zhu ; Xiaofeng Ling
- [127]. IEEE Access
- [128]. Year: 2020 | Volume: 8 | Journal Article <https://ieeexplore.ieee.org/document/8993810/>
- [129]. Improved stochastic modeling of shapes for content-based image retrieval
- [130]. S. Muller ; G. Rigoll
- [131]. Proceedings IEEE Workshop on Content-Based Access of Image and
- [132]. Video Libraries (CBAIVL'99) Year: 1999 | Conference Paper <https://ieeexplore.ieee.org/document/781118/>
- [133]. [32] Defining image content with multiple regions-of-interest
- [134]. B. Moghaddam ; H. Biermann ; D. Margaritis
- [135]. Proceedings IEEE Workshop on Content-Based Access of Image and
- [136]. Video Libraries (CBAIVL'99)
- [137]. Year: 1999 | Conference Paper
- [138]. Cited by: Papers (27) | Patents (11) <https://ieeexplore.ieee.org/document/781130/>
- [139]. [33] Image background search: combining object detection techniques with content-based image retrieval (CBIR) systems
- [140]. R. Srihari ; Zhongfei Zhang ; Aibing Rao
- [141]. Proceedings IEEE Workshop on Content-Based Access of Image and
- [142]. Video Libraries (CBAIVL'99)
- [143]. Year: 1999 | Conference Paper Cited by: Papers (6) | Patents (6) <https://ieeexplore.ieee.org/document/781131/>