EmoSense

Music mood Detector & Recommender System

Group - 55

Index

Project Description

This project recommends music to users based on whatever mood they want to listen. The system takes in music data in form of audio, video and text and performs sentiment analysis on them to find out the mood in that music. Then using a clustering and other recommendation techniques to recommend music based on the mood wanted by the user.

Motivation

Current music classification techniques do not take into account genre fusions, range of emotions and try to classify music under one category. Leading to annoying recommendations filling up your feed. This project will make use of cutting-edge ML , NLP and CV methods to accurately categorise music into different emotions expressed based on its audio, video, and lyrical content. Hence, improving the quality of recommendation systems.

Data & Attributes

- Relevant Datasets- millionsong datasets
- Spotify api to extract lyrics of songs.
- Spotify api - audio data as an array of loudness , beats etc.
- Raw audio data can also be used and feeded into a sequential network along with lyrics.
- Video data of music videos can also be used to provide better context during sentiment analysis.

Input Type 1

- Input text from User- keywords or phrases
- Speech recognition- convert user's voice input to text.
- Analysing user's previous music choices- extracting the audio features of the music tracks
- User Activity- search history, playlist creation, and playback history.
- Classifying the emotion of a music video- using Lyrical Text, background track and Video features.

Input type 2

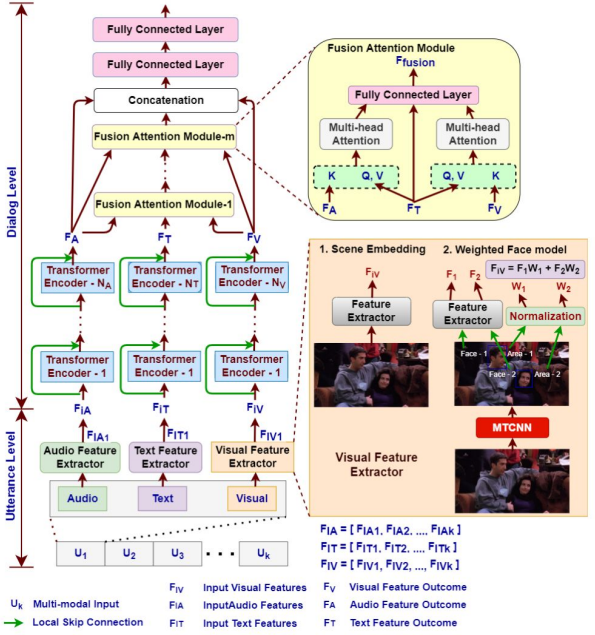
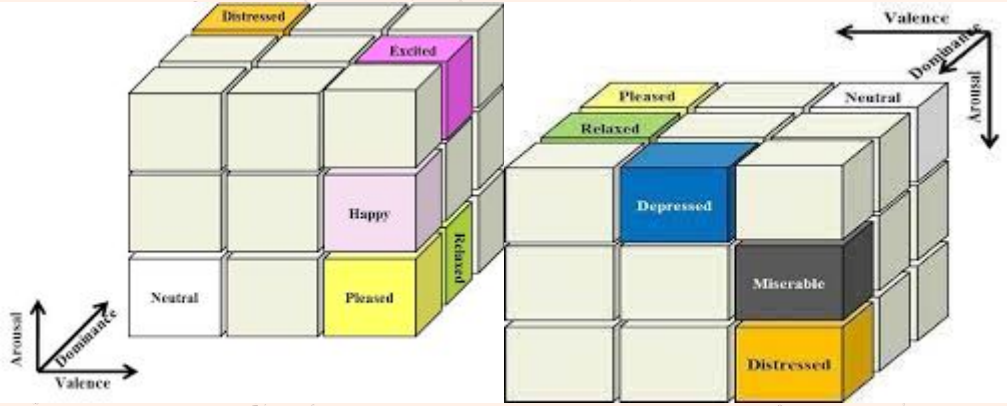


Figure 3. Network design of the proposed framework.

- Video input- divided into frames at frequency (n Hz).
- Single embedding computed for each frame (image + text) (Akbari et al., 2021)
- Since the emotion value for each frame is not available, the embedding will be useful for labeling the image data.

Output Type 1



One or more model will take in lyrics, audio and video data as input and output 3 values -

- **Dominance -** degree of control exerted by a stimulus
- **Valence** - pleasantness of a stimulus
- **Arousal -** intensity of emotion provoked by a stimulus

These Values are 3 components of emotions or basically how emotion is quantified

https://ieeexplore.ieee.org/stamp/stamp.jsp?arnumber=9060904

Output Type 2

- Build a recommendation system based on the emotions derived.
- The recommendations could provide a preview of the songs that match user's mood.
- The user could also create their own playlist from recommended songs
- Additionally, the system can offer genre-classified selections to help the user choose the music they want to listen to.

Methodology

- Data collection
- Data preprocessing
- Feature extraction
- Classifying mood using ML algos
- Recommendation generation
- Performance evaluation

Social Impact

- **Improved mental health** - Listening to music can have a positive impact on mental health, and a mood-based music recommendation system can help users find music that is appropriate for their mood
- **Increased social connectedness-** Music is often a shared experience, and a mood-based music recommendation system could help users connect with others who have similar music tastes or moods.
- **Cultural and artistic exposure-** A mood-based music recommendation system could introduce users to new artists, genres, and cultures that they may not have otherwise been exposed to

Fusing audio, visual & textual clues for sentiment analysis from multimodal content

Soujanya Poria [a], Erik Cambria [b], Newton Howard [c], Guang-Bin Huang [d], Amir Hussain [a]

- Did Sentiment analysis of Youtube Dataset
- Video Data ➡ Images , Audio and text data
- Vectors Representation
- Fused together vectors
- ELM , Naive Bayes , Neural Networks , SVM.

They achieves an accuracy of nearly 80%, outperforming all state-of-the-art systems by more than 20%.

Automatic mood detection & tracking of music audio signals

Lie Lu; D. Liu; Hong-Jiang Zhang

- Extracted features of intensity (sound level), timbre (spectrum), and rhythm (tempo).
- Intensity features and timbre features are extracted from each frame.
- Rhythm features are extracted at the level of the whole music clip
- 4 output features - Contentment, Depression, Exuberance and Anxious/Frantic
- Model - Gaussian mixture model
- Achieved high accuracy.

Novelty

- Multimodal Sentiment Analysis is emerging approach in Sentiment analysis
- Most older models used only text data
- Music can have words that mean something else but actually have opposite meanings/emotions.
- Nobody has explored using all three modalities in only music data.

References

1. Akbari, H., Yuan, L., Qian, R., Chuang, W.-H., Chang, S.-F., Cui, Y., & Gong, B. (2021). VATT: Transformers for Multimodal Self-Supervised Learning from Raw Video, Audio and Text. Advances in Neural Information Processing Systems, 34, 24206–24221. https://proceedings.neurips.cc/paper/2021/hash/cb3213ada48302953cb0f166464ab356-Abstract.html
2. Bhat, A. S., Amith, V. S., Prasad, N. S., & Mohan, D. M. (2014). An Efficient Classification Algorithm for Music Mood Detection in Western and Hindi Music Using Audio Feature Extraction. 2014 Fifth International Conference on Signal and Image Processing, 359–364. https://doi.org/10.1109/ICSIP.2014.63
3. Bogdanov, D., Lizarraga Seijas, X., Alonso-Jiménez, P., & Serra, X. (2022). MUSAV: A dataset of relative arousal-valence annotations for validation of audio models. http://repositori.upf.edu/handle/10230/54181
4. Chudasama, V., Kar, P., Gudmalwar, A., Shah, N., Wasnik, P., & Onoe, N. (2022). M2FNet: Multi-modal Fusion Network for Emotion Recognition in Conversation (arXiv:2206.02187). arXiv. https://doi.org/10.48550/arXiv.2206.02187
5. datancoffee. (2017, May 4). Opinion Analysis of Text using Plutchik – Data and Coffee. https://datancoff.ee/2017/05/opinion-analysis-of-text-using-plutchik/
6. DEAM dataset—Database for Emotional Analysis of Music. (n.d.). Retrieved February 21, 2023, from https://cvml.unige.ch/databases/DEAM/
7. Delbouys, R., Hennequin, R., Piccoli, F., Royo-Letelier, J., & Moussallam, M. (2018). Music Mood Detection Based On Audio And Lyrics With Deep Neural Net (arXiv:1809.07276). arXiv. https://doi.org/10.48550/arXiv.1809.07276
8. Doshi, K. (2021, May 21). Audio Deep Learning Made Simple: Sound Classification, step-by-step. Medium. https://towardsdatascience.com/audio-deep-learning-made-simple-sound-classification-step-by-step-cebc936bbe5
9. Edmonds, D., & Sedoc, J. (2021). Multi-Emotion Classification for Song Lyrics. Proceedings of the Eleventh Workshop on Computational Approaches to Subjectivity, Sentiment and Social Media Analysis, 221–235. https://aclanthology.org/2021.wassa-1.24
10. Han, D., Kong, Y., Han, J., & Wang, G. (2022). A survey of music emotion recognition. Frontiers of Computer Science, 16(6), 166335. https://doi.org/10.1007/s11704-021-0569-4
11. Hashemi, A. (2022, January 17). Unsupervised Semantic Sentiment Analysis of IMDB Reviews. Medium. https://towardsdatascience.com/unsupervised-semantic-sentiment-analysis-of-imdb-reviews-2c5f520fbf81
12. Huddar, M., Sannakki, S., & Rajpurohit, V. (2019). A Survey of Computational Approaches and Challenges in Multimodal Sentiment Analysis. INTERNATIONAL JOURNAL OF COMPUTER SCIENCES AND ENGINEERING, 7, 876–883. https://doi.org/10.26438/ijcse/v7i1.876883
13. Jia, X. (2022). A Music Emotion Classification Model Based on the Improved Convolutional Neural Network. Computational Intelligence and Neuroscience, 2022, 6749622. https://doi.org/10.1155/2022/6749622
14. Jia, Y. (2020). A Deep Learning System for Sentiment Analysis of Service Calls. Proceedings of the 3rd Workshop on E-Commerce and NLP, 24–34. https://doi.org/10.18653/v1/2020.ecnlp-1.4
15. Kaur, R., & Kautish, S. (2022). Multimodal Sentiment Analysis: A Survey and Comparison [Chapter]. Research Anthology on Implementing Sentiment Analysis Across Multiple Disciplines; IGI Global. https://doi.org/10.4018/978-1-6684-6303-1.ch098
16. Koromilas, P., & Giannakopoulos, T. (2022). Unsupervised Multimodal Language Representations using Convolutional Autoencoders (arXiv:2110.03007). arXiv. https://doi.org/10.48550/arXiv.2110.03007
17. Laurier, C., Grivolla, J., & Herrera, P. (2008). Multimodal Music Mood Classification Using Audio and Lyrics. 688–693. https://doi.org/10.1109/ICMLA.2008.96
18. Lerch, A. (2023). Mood Recognition. In An Introduction to Audio Content Analysis: Music Information Retrieval Tasks and Applications (pp. 337–346). IEEE. https://doi.org/10.1002/9781119890980.ch13
19. Liu, G., & Tan, Z. (2020). Research on Multi-modal Music Emotion Classification Based on Audio and Lyirc. 2020 IEEE 4th Information Technology, Networking, Electronic and Automation Control Conference (ITNEC), 1, 2331–2335. https://doi.org/10.1109/ITNEC48623.2020.9084846
20. Lu, L., Liu, D., & Zhang, H.-J. (2006). Automatic mood detection and tracking of music audio signals. IEEE Transactions on Audio, Speech, and Language Processing, 14(1), 5–18. https://doi.org/10.1109/TSA.2005.860344
21. Meyers, O. C. (2007). A mood-based music classification and exploration system [Thesis, Massachusetts Institute of Technology]. https://dspace.mit.edu/handle/1721.1/39337
22. Pandeya, Y. R., Bhattarai, B., & Lee, J. (2021). Music video emotion classification using slow–fast audio–video network and unsupervised feature representation. Scientific Reports, 11(1), Article 1. https://doi.org/10.1038/s41598-021-98856-2
23. Pandeya, Y. R., & Lee, J. (2021). Deep learning-based late fusion of multimodal information for emotion classification of music video. Multimedia Tools and Applications, 80(2), 2887–2905. https://doi.org/10.1007/s11042-020-08836-3

THANK YOU !