# Berkeley SCHOOL OF INFORMATION

# Sentiment Analysis of Songs Lyrics

*An NLP approach using deep learning*

David Gamez, Joseph Mercer
*david.gamez@berkeley.edu, jlmercer@berkeley.edu*

UC Berkeley School of Information. Master of Information and Data Science.
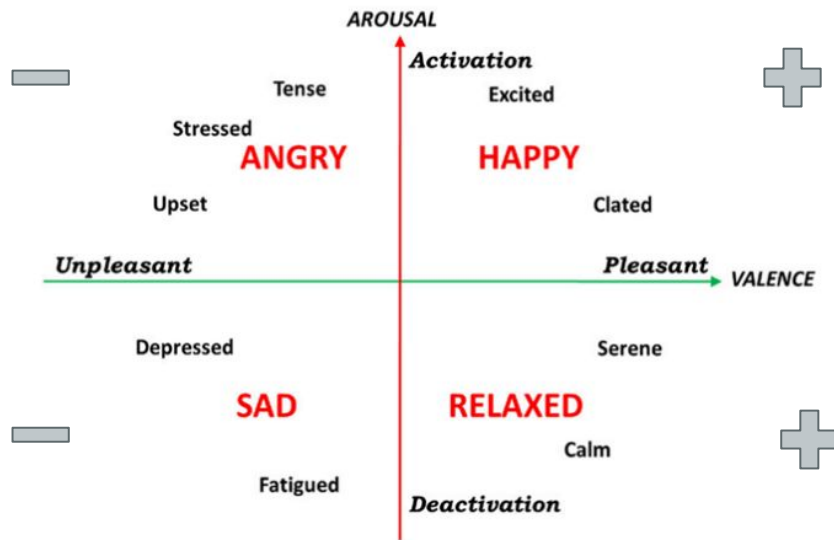W266 Natural Language Processing with Deep Learning

# Objectives

**Binomial classification**

Positive/Negative valence

**Multinomial classification:**

Four emotions categories of Russell's model: happy, angry, sad and relaxed

# Data

Two datasets:

- corpus of collection of 771 song lyrics collected from Allmusic, annotated using human evaluators and user tags
- collection of 2,500 song lyrics annotated using content words of lyrics and their valence and arousal norms in affect lexicons only
- Data pre-processing (spelling errors / repeated syllables)
- Final Corpus: 2574 songs

| Mood | Valence | Source 1 [4] | Source 2 [1] | Total |
|---|---|---|---|---|
| Happy | Positive | 211 | 465 | 667 |
| Angry | Negative | 204 | 468 | 672 |
| Sad | Negative | 204 | 443 | 647 |
| Relaxed | Positive | 148 | 440 | 588 |

**2574 songs!**

# Binomial

**Best result:**

BERT model on accuracy and f-1 score on test data

**Interesting:**

DL models get more homogeneous results on precision and recall for both classes

| Model | Metrics | | | | |
|---|---|---|---|---|---|
| | class | precision | recall | f1-score | # |
| Baseline | Accuracy: **0.48** | | | | 515 |
| NB | pos | **0.88** | **0.61** | **0.72** | 265 |
| | neg | **0.69** | **0.91** | 0.78 | 250 |
| | Accuracy: 0.7553 | | | | 515 |
| CNN | pos | **0.65** | **0.82** | 0.72 | 155 |
| | neg | 0.78 | **0.59** | 0.78 | 169 |
| | Accuracy: **0.7006** | | | | 324 |
| Bi-LSTM RNN | pos | 0.72 | 0.72 | 0.72 | 155 |
| | neg | 0.74 | 0.74 | **0.74** | 169 |
| | Accuracy: 0.7315 | | | | 324 |
| BERT | pos | 0.80 | 0.75 | **0.76** | 256 |
| | neg | **0.795** | 0.77 | **0.78** | 259 |
| | Accuracy: **0.7825** | | | | 515 |

# Multinomial

**Best result:**

Naive Bayes model on accuracy and most of f-1 score on test data

**Interesting:**

DL models get more homogeneous results on precision and recall for both classes

| Model | Metrics | | | | |
|---|---|---|---|---|---|
| | class | precision | recall | f1-score | # |
| Baseline | Accuracy: **0.2563** | | | | 515 |
| NB | happy | 0.59 | 0.74 | **0.66** | 132 |
| | angry | 0.84 | 0.72 | **0.77** | 143 |
| | sad | 0.51 | 0.66 | **0.58** | 122 |
| | relax | 0.57 | 0.33 | 0.42 | 118 |
| | Accuracy: **0.6233** | | | | 515 |
| CNN | happy | 0.60 | 0.62 | 0.61 | 90 |
| | angry | 0.63 | 0.79 | 0.70 | 85 |
| | sad | 0.46 | 0.46 | 0.46 | 70 |
| | relax | 0.60 | 0.42 | 0.49 | 79 |
| | Accuracy: 0.5802 | | | | 324 |
| Bi-LSTM RNN | happy | 0.53 | 0.57 | **0.55** | 90 |
| | angry | 0.68 | 0.66 | **0.67** | 85 |
| | sad | 0.35 | 0.34 | **0.35** | 70 |
| | relax | 0.52 | 0.51 | **0.51** | 79 |
| | Accuracy: **0.5277** | | | | 324 |

# Multinomial

**Normalized confusion matrix for Bi-LSTM RNN model**

| True\Pred | Happy | Angry | Sad | Relaxed |
|-----------|-------|-------|-----|---------|
| Happy | **56.67** | 14.44 | 13.33 | 15.56 |
| Angry | 10.59 | **65.88** | 17.65 | 5.88 |
| Sad | 25.71 | 14.29 | **34.29** | 25.71 |
| Relaxed | 22.78 | 3.80 | 22.78 | **50.63** |

**Interesting:**

Angry songs are the easiest to classify: very frequent use of rude vocabulary and expressions, not common in any of the other classes. However, the boundaries between a sad and relaxed, or between a relaxed and happy song lyrics are not always so clear, even for a human being

# Conclusion

- For binomial classification, we developed a deep learning model which was able to beat the classical model, which it is a better result than the one reported in reference #2, where a combination of lyrics and audio models was required to beat a classical NLP approach -on lyrics only-.

- On top of that, our BERT model got a **78.25%** accuracy on positive/negative (valence), beating the 76.29% maximum accuracy on valence reported in reference #6 and even the 77.23% reported on arousal, which we did not train for.

- However, for multiclass classification, we were not able to beat a classical model.

- The greatest confusion of the model was found between relaxed and sad emotions. A larger corpus would probably help dealing with this difficulty.

# References

[1]     Çano, Erion; Morisio, Maurizio (2017). MoodyLyrics: A Sentiment Annotated Lyrics Dataset. In: 2017 International Conference on Intelligent Systems, Metaheuristics & Swarm Intelligence, Hong Kong, March, 2017. pp. 118-124 (Link [1]).

[2]     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 e-prints arXiv:1809.07276 (Link [2]).

[3]     Kim, Yoon (2014). Convolutional Neural Networks for Sentence Classification. Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing. Association for Computational Linguistics (Link [3]).

[4]     Malheiro, Ricardo & Panda, Renato & Gomes, Paulo & Paiva, Rui Pedro. (2016). Classification and Regression of Music Lyrics: Emotionally-Significant Features. 45-55. 10.5220/0006037400450055 (Link [4]).

[5]     Russell, James A.(1980). A circumplex model of a_ect. Journal of personality and social psychology, 39(6):1161-1178, 1980 (Link [5]).

[6]     Zaanen, Kanters (2010). Automatic mood classification using tf*idf based on lyrics. Proceedings of the 11th International Society for Music Retrieval Conference, pages 75-80 (Link [6])

Thanks for listening!