Design and Implementation of Affective Android Music Player

Michal Staszewski

Kongens Lyngby 2012
IMM-MSc-2012-94

Technical University of Denmark
Informatics and Mathematical Modelling
Building 321, DK-2800 Kongens Lyngby, Denmark
Phone +45 45253351, Fax +45 45882673
reception@imm.dtu.dk
www.imm.dtu.dk IMM-MSc-2012-94

Summary (English)

An affective load estimation model based on extended Affective Norms for English Words list [2] has been created. The model is used to assess emotional load (in terms of valence and arousal) of a corpus of lyrics. The ANEW [2] list has been extended with an algorithm using Latent Semantic Analysis [3] to compute term to term similarities which were then used to derive affective estimations for words not in the ANEW [2] list. Correlation between lyrics affective load and audio features is confirmed using Pearson measure to detect statistical significance. An Android music player is subsequently developed and audio parameters and lyrics affective load are used by the algorithm implemented in this application to assign a point in valence/arousal space for selected songs. The application allows the user to change assigned valence/arousal parameters. Bayesian classifier based model is trained to predict user preferences. An affective load estimation model based on extended Affective Norms for English Words list [1] has been created. The model is used to assess emotional load (in terms of valence and arousal) of a corpus of lyrics. The ANEW [1] list has been extended with an algorithm using Latent Semantic Analysis [2] to compute term to term similarities which were then used to derive affective estimations for words not in the ANEW [1] list. Correlation between lyrics affective load and audio features is confirmed using Pearson measure to detect statistical significance. An Android music player is subsequently developed and audio parameters and lyrics affective load are used by the algorithm implemented in this application to assign a point in valence/arousal space for selected songs. The application allows the user to change assigned valence/arousal parameters. Bayesian classifier based model is trained to predict user preferences.

Contents

Summary (English) **i**

1 Introduction and Motivation **1**

2 Problem Definition **3**

3 Related Work **5**

4 Data Aquisition **7**

5 Measuring Affective Load **9**

6 Extending ANEW **11**
6.0.1 Affective load and audio features 14
6.0.2 Fine - grained analysis 15

7 Predicting Users Preferences **17**

8 Discussion and Conclusions **19**

Bibliography **22**

Chapter 1

Introduction and Motivation

It is postulated that if correlation between the affective load of lyrics and the audio features exists, not only the audio features, but also the lyrics can be used to produce more precise playlist suggestions in a variety of applications such as online radios and mobile music players. The proposed method of extraction of affective load data from lyrics transformed into a bag-of-words model [3] using the LSA [4] – extended Affective Norms for English Words [5] is not resource demanding and can be performed on a mobile device most likely faster than in-depth audio analysis. Affective data can be used to facilitate playlist suggestions, possibly extending systems such as PATS and PlaylistDJ discussed in [6] by serving as an additional music describing parameter.

Chapter 2

Problem Definition

The aim of this thesis is to find out if there is a relation between the proposed affective model and the EchoNest [2] audio parameters like energy, tempo and loudness. Pearson correlation will be employed to determine statistical significance of findings. The proposed affective model is described in the 'Measuring Affective Load' chapter.

The model will be then employed to build a music player able to predict user preferences.

Chapter 3

Related Work

K. K. Koclega [7] has performed analysis where a small sample of 32 songs has been used. LSA has been used to find similarities between lines of lyrics and emotional tags such as 'happy', 'sad', 'mellow'. There have been 12 tags used in total. Cosine similarities between each tag, a line of lyrics and the whole song have been calculated. One of the issues with this approach was frequent lack of occurrence of a tag ('query' in LSA [4] terms) and the song's lyrics ('document'). Despite those problems correlation between such emotional tags as 'happy' and 'sad', 'romantic' and 'melancholy' has been found. In this thesis a different approach is taken. There is a simple but well established 2D affective model of arousal and valence [8] postulating that a temporal emotion can be roughly described as a point on arousal v. valence plane. Affective Norms for English Words dataset is using this model and provides valence and arousal values for 1034 words. The context in which a word is used is vitally important for its meaning and co-occurring words have similar or associated meanings [4]. LSA can be employed to measure co-occurrence of terms in a set of documents. This is the base for the extended-ANEW model, where if a term is similar enough to an already existing ANEW term, it is included in the model with affective values similar to the original term's values weighted by LSA's cosine similarity measure. Then the lyric's bag-of-words affective load is evaluated using the extended ANEW list both on whole song and line of lyrics scale. The emergent affective load is analysed together with the EchoNest obtained music data.

Chapter 4

Data Aquisition

MusixMatch list of EchoNest track ids, artist names and song titles has been used as a base for analysis because it claims to contain less purely instrumental tracks and less duplicates [9] [10]. There are ca. 770 000 track entries in the list. Lyrics has been obtained for ca. 207 000 songs from the LyricWiki [11] using a crawler which would process only ASCII containing lyrics. This is by design, it happened to be the most efficient way of filtering non- English lyrics tried, that is better than both letter frequency testing and dictionary testing method, ASCII-filtering will however remove a number of songs with English lyrics, but the final number of above 207 000 seems good for further analysis. Afterwards each song has been evaluated for valence, arousal and dominance values by transforming the lyrics into a bag of words model, selecting 20 most commonly occurring words (excluding those from a stoplist), and taking the average ANEW affective values (valence, arousal and dominance) of those most frequent 20 words. This created a set of 132 000 emotion-annotated songs.

The number is less than before because of the filtering strategy. It has been assumed that song's lyrics often contain onomatopoeic word repetitions. Furthermore ASCII – filtering have not removed all non- English songs. The mentioned frequency distribution has been employed to address the first issue. That is, if there is an onomatopoeic word 'la' occurring 10x more often than all the other words and it is present in the ANEW set, then the average over all words would over represent it, whereas it would appear in the top 20 words list only once.

The next step was to select 20 000 lowest and highest valence songs. This is because of a postulate that if there is a correlation between lyrics' valence (or arousal) and audio parameters such as loudness or energy it will be the most apparent in the extremes of the affective range. The last step was to obtain EchoNest audio analysis data for the 20 000 songs. pyEchoNest (11) API has been employed to perform this task.

Chapter 5

Measuring Affective Load

Semantic Analysis works as follows:

- Term occurrences in each document (where document any text fragment, length of which depends on how fine or coarse grained co-occurrence patterns we would like to find, how much 'noise' there is in the data. It is commonly a single article, paragraph or line) are saved in a term by document matrix where rows represent terms and columns documents. For each occurrence of a term in a document the matrix value is increased by a constant (e.g. 1).
- Weights of term occurrences are calculated by tf-idf. A high weight in tf-idf is reached by a high term frequency (in the given document) and a low document frequency of the term in the whole collection of documents. Tf-idf matrix element is defined as: $|D| / noOfOccurences(t)$.
- Next step is the Singular Value Decomposition. It's defined as such an operation (decomposition) which transforms matrix X into matrices $X = U\Sigma V^T$ where U and V are orthogonal and Σ is diagonal. It turns out that selecting the k largest diagonal values of Σ, and their corresponding singular vectors from U and V, and performing the dot product gives rank k approximation to X with the minimal error.

Chapter 6

Extending ANEW

For each ANEW word the most similar words obtained in the way described have been selected. The threshold similarity value for inclusion has been 0.7 in cosine similarity. Extended ANEW sets have been prepared. Following is an example from the Brown corpus extended ANEW with 100 factors: barge 4.51423824 4.52226736 4.598544 weary 0.700728 soaked 4.50831166 4.51643874 4.593646 weary 0.703177 shocking 4.48527326 4.49378114 4.574606 weary 0.712697 There is no aparent meaning in those relations, but when statistical methods are applied, one can potentially see a correlation on a large corpora [1]

Computing Affective Load First the plain ANEW and bag-of-words approach has been utilized to obtain a list of 40 000 lowest and highest valence songs as described in the 'Data Acquisition' chapter. Next a list of 400 (200 low and 200 high valence) songs with the best plain ANEW word to overall number of words ratio has been selected. This list has been analysed further to obtain Pearson's correlation values of arousal and valence, because of the assumption that a good model would show strong correlation between the two [8]. As a result ANEW extended with Brown corpus with 100 LSA topics has been chosen. Furthermore Porter stemmer has been used on the song's lyrics and the extended ANEW wordlist to include more lyrics terms in the analysis.

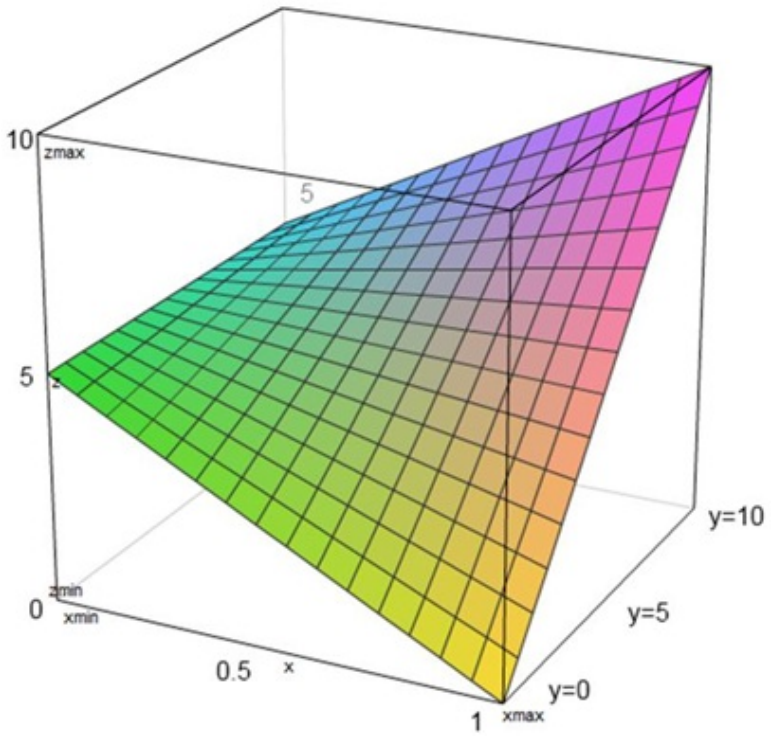

Figure 6.1: Function computing substituted values for emotional load based on cosine similarity in LSA space between the original and substitute term chart where z is the resultant affective value, y is original value and x is the similarity measure

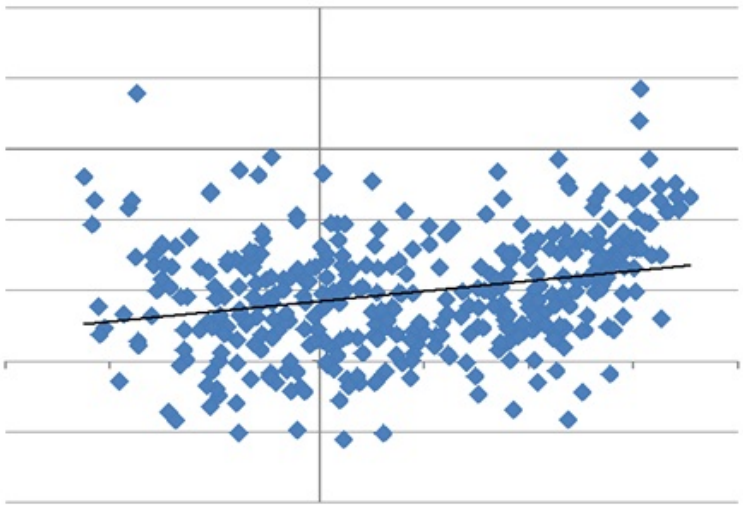

Figure 6.2: the Arousal/Valence chart has been shifted by [-5,-5] to show the emotional plane more clearly. Lower left part (low valence and low arousal) is classified as sad and depressed, upper left as angry, lower right as mellow, soft and upper right as happy (Brown corpus, 100 LSA factors)

One tailed Pearson test with hypotheses p = 0 and p != 0 (there is or there is no correlation) with the confidence level of 5% and df = 393 - 2 has been employed on a sample of 393 songs. The test is passed for Brown corpus extended LSA since p = 0.5045 and for df = 391 the test is passed for p > 0.113. This list has been analysed further to obtain Pearson's correlation values of arousal and valence, because of the assumption that a good model would show strong correlation between the two [8]. As a result ANEW extended with Brown corpus with 100 LSA topics has been chosen. Furthermore Porter stemmer has been used on the song's lyrics and the extended ANEW wordlist to include more lyrics terms in the analysis. The least squares method for a linear regression of arousal and valence relation gives the following results: *Dataset R*

ANEW	0.0674
Brown100	0.2546
Brown200	0.0623
Brown400	0.1349
Wiki100	0.0108
Wiki200	0.0137
Wiki400	0.0429

The above means that the linear regression of arousal and valence explains around 25% of data variance for the 'Brown100' list. The next step is to find a correlation between EchoNest energy (exact definition not available, but includes dynamic range, rhythmic regularity, timbral components, tempo, and other timing features (...)' [2], tempo (precise definition not available, unofficial sources tell it is synonymous with the Beats Per Minute measure) and loudness (average loudness in dB for the whole song).

6.0.1 Affective load and audio features

There is a statistically significant correlation between the valence and energy (p = -0.17, α= 0.05) and valence and loudness (p = -0.18, α=0.05) for the analysed sample. Statistically significant correlation between the tempo and valence, tempo and arousal, energy and arousal and loudness and arousal has not been found.

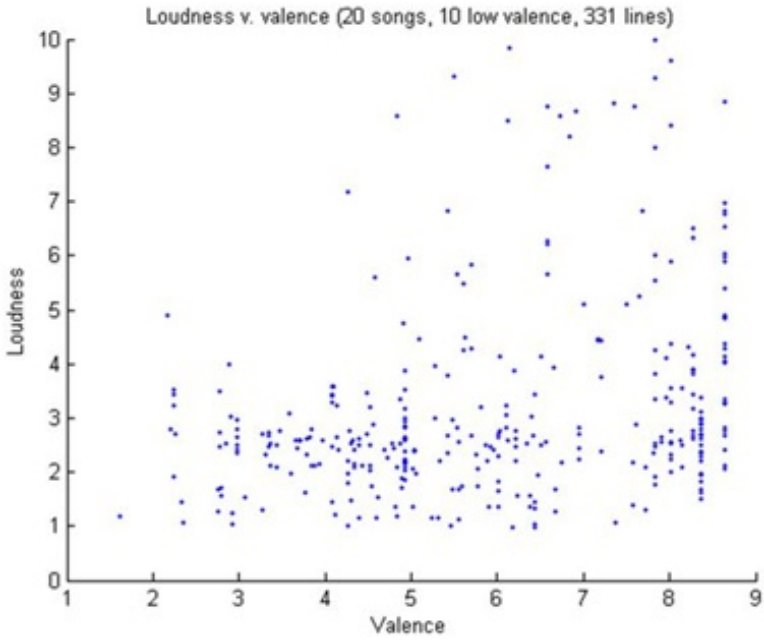

Figure 6.3: each point represents a line of lyrics. Loudness has been inverted and scaled to 0-10 range. It is visible that the lines of higher valence are louder.

6.0.2 Fine - grained analysis

10 randomly selected high valence and 10 randomly selected low valence songs lines of lyrics have been analysed using the bag-of-words model with Brown extended ANEW (100 factors) and Porter's stemmer. If the affective load of a line of lyrics was undetermined (word not in the extended list) then the line and its' corresponding audio data (average loudness) has been removed. It was assumed that the lyrics and segments of audio data are linearly distributed in time. Valence and loudness correlation analysis shows significant correlation (p = -0.3, df = 329) on a line of lyrics level for the given sample.

Chapter 7

Predicting Users Preferences

Social metadata has become an important aid in content classification. [15] Bayesian methods of content assesment are widely used in spam filters and it is also well established that these methods allow for classification of patterns of media consumption. This can help discover connections among the modern social network peers and facilitate knowledge transfer. The idea behind the Affective Media Player was to make it able to recognize emotion and adapt to them by adjusting the music suggestions. What has been delivered is a proof of concept. It performs measurably better with the 'Extended ANEW' word list, the area of further study should be much about the most reliable way of predicting personal preferences in music. I believe there is much more in this field to have a closer look at. The Bayesian model as it is now is able to show which words from the lyrics are the most likely to be associated with the same kind of music the user would prefer, but this, of course, is just the beginning.

Chapter 8

Discussion and Conclusions

Statistical methods employed are limited to linear dependence. While it works well for arousal/valence model it might not be the best option for some of the audio features, such as tempo. Proposed 'extended ANEW' model works in practise, but the function of translating cosine similarity into emotional values is invented based only on the real-life results and not psychological modelling.

Bibliography

[1] Landauer, T. K., Foltz, P. W., and Laham, D. (1998). *Introduction to Latent Semantic Analysis.* Discourse Processes, 25, 259-284.

[2] *EchoNest Platform.* [Online] [Cited: 15 08 2012.] `http://the.echonest.com/platform/`.

[3] *Bag of words model.* Wikipedia. [Online] [Cited: 20 08 2011.] `http://en.wikipedia.org/wiki/Bag_of_words_model`.

[4] Scott Deerwester, Susan T. Dumais, George W. Furnas, Thomas K. Landauer, Richard Harshman. *Indexing by Latent Semantic Analysis.* s.l. : Journal of the American Society for Information Science. 41(6):391-407, 1990.

[5] Bradley, M.M., and Lang, P.J. *Affective Norms for English Words (ANEW): Instruction Manual and Affective Ratings.* s.l. : The Center of Research in Psychophysiology, University of Florida, 1999.

[6] Handler, Rasmus and Zandi, Nima. *People, Places and Playlists.* 2800 Kongens Lyngby, Denmark : Technical University of Denmark, 2010.

[7] Koclega, Kamil Krzysztof. *Cognitive semantics of lyrics and audio features in songs.* Kongens Lyngby : s.n., 2010.

[8] PA Lewis, HD Critchley, P Rotshtein and RJ Dolan. *Neural Correlates of Processing Valence and Arousal in Affective Words.* s.l. : Oxford University Press, 2006.

[9] *musixmatch.* [Online] `http://labrosa.ee.columbia.edu/millionsong/musixmatch`.

[10] *The Million Song Dataset.* Thierry Bertin-Mahieux, Daniel P.W. Ellis, Brian Whitman, Paul Lamere. 2011. Proceedings of the 12th International Conference on Music Information Retrieval (ISMIR).

[11] *LyricWiki.* [Online] [Cited: 01 07 2011.] `http://lyrics.wikia.com/`.

[12] *pyEchoNest.* [Online] `http://code.google.com/p/pyechonest/`.

[13] . *Latent Semantic Analysis.* Wikipedia. [Online] `http://en.wikipedia.org/wiki/Latent_semantic_analysis`.

[14] W. N. Francis, H. Kucera. *A Standard Corpus of Present-Day Edited American English for use with Digital Computers.* Providence : Department of Linguistics, Brown University, 1964.

[15] Jude Yewa and David A. Shammaa *Know Your Data: Understanding Implicit Usage versus Explicit Action in Video Content Classiffication* a Yahoo! Research, 4301 Great America Parkway, Santa Clara, USA;