Inferring the meaning of chord sequences via lyrics

Tom O'Hara
Computer Science Department
Texas State University
San Marcos, TX
to17@txstate.edu

ABSTRACT

This paper discusses how meanings associated with chord sequences can be inferred from word associations based on lyrics. The approach works by analyzing in-line chord annotations of lyrics to maintain co-occurrence statistics for chords and lyrics. This is analogous to the way parallel corpora are analyzed in order to infer translation lexicons. The result can benefit musical discovery systems by modeling how the chord structure complements the lyrics.

Categories and Subject Descriptors

H.5.5 [**Sound and Music Computing**]: Modeling

General Terms

Experimentation

Keywords

Music information retrieval, Natural language processing

1. INTRODUCTION

A key task for music recommendation systems is to determine whether an arbitrary song might match the mood of the listener. An approach commonly used is for a system to learn a classification model based on tagged data (i.e., supervised classification). For example, training data might be prepared by collecting a large variety of songs and then asking users to assign one or more mood categories to each song. Based on these annotations, a model can be developed to assign the most likely mood type for a song, given features derived from the audio and lyrics.

Such an approach works well for capturing the mood or other meaning aspects of entire songs, but it is less suitable for capturing similar aspects for segments of songs. The main problem is that human annotations are generally only done for entire songs. However, for complex songs this might lead to improper associations being learned (e.g., a sad introduction being tagged upbeat in a song that is otherwise upbeat). Although it would be possible for segments to be annotated as well, it would not be feasible. There would simply be too many segments to annotate. Furthermore, as the segments get smaller, the annotations would become more subjective (i.e., less consistent). However, by using lyrics in place of tagged data, learning could indeed be done at the song segment level.

Parallel text corpora were developed primarily to serve multilingual populations but have proved invaluable for inducing lexicons for machine translation [6]. Similarly, a type of resource intended for musicians can be exploited to associate meaning with music. Guitarists learning new songs often rely upon tablature notation ("tabs") provided by others to show the finger placement for a song measure by measure. Tabs often include lyrics, enabling note sequences to be associated with words. They also might indicate chords as an aid to learning the sequence (as is often done in scores for folk songs). In some cases, the chord annotations for lyrics are sufficient for playing certain songs, such as those with accompaniment provided primarily by guitar strumming.

There are several web sites with large collections of tabs and chord annotations for songs (e.g., about 250,000 via www.chordie.com). These build upon earlier Usenet-based guitar forums (e.g., alt.guitar.tabs). Such repositories provide a practical means to implement unsupervised learning of the meaning of chord sequences from lyrics. As these resources are willingly maintained by thousands of guitarists and other musicians, a system based on them can be readily kept current. This paper discusses how such resources can be utilized for associating meaning with chords.

2. BACKGROUND

There has been a variety of work in music information retrieval on learning the meaning of music. Most approaches have used supervised classification in which user tags serve as ground truth for machine learning algorithms. A few have inferred the labels based on existing resources. The approaches differ mainly on the types of features used. Whitman and Ellis [10] combine audio features based on signal processing with features based on significant terms extracted from reviews for the album in question, thus an unsupervised approach relying only upon metadata about songs (e.g., author and title). Turnbull et al. [9] use similar types of audio features, but they incorporate tagged data describing the song in terms of genre, instrumentality, mood, and other attributes. Hu et al. [2] combine word-level lyrics and audio features, using tags derived from social media, filtered based on degree of affect, and then revised by humans (i.e., partly

WOMRAD 2011 2nd Workshop on Music Recommendation and Discovery, colocated with ACM RecSys 2011 (Chicago, US)

1. Obtain large collection of lyrics with chord annotations
2. Extract lyrics proper with annotations from dataset
3. *Optional:* Map lyrics from words to meaning categories
   (a) Get tagged data on meaning categories for lyrics
   (b) Preprocess lyrics and untagged chord annotations
   (c) Train to categorize over words and hypernyms
   (d) Classify each lyric line from chord annotations
4. Fill contingency table with chord(s)/token associations
5. Determine significant chord(s)/token associations.

Figure 1: Process in learning meanings for chord sequences. The meaning *token* is either an individual word or a meaning category label; and, *chord(s)* can be a single chord or a four-chord sequence.

supervised). McKay at al. [5] combine class-level lyric features (e.g., part of speech frequencies and readability level) with ones extracted from user tags from social media (specifically Last.fm[1]) as well as with features derived from general term co-occurrence via web searches for the task of genre classification.

Parallel corpora are vital for machine translation. Fung and Church [1] induce translation lexicons by tabulating co-occurrence statistics over fixed-size blocks, from which contingency tables are produced to derive mutual information statistics. Melamed [6] improves upon similar approaches by using a heuristic to avoid redundant links.

3. PROCESS

The overall task of processing is as follows: starting with a large collection of lyrics with chord annotations, infer meaning category labels for the chord sequences that occur, based on word associations for the chords sequences. Several steps are required to achieve this in order to make the lyrics more tractable for processing and due to the option for including a lyrics classifier as a refinement of the main induction step. The latter allows meaning to be in terms of high-level mood categories rather than just words.

Figure 1 lists the steps involved. First the Internet is checked to find and download a large sample of lyrics with word annotations. The resulting data then is passed through a filter to remove extraneous text associated with the lyrics (e.g., transcriber notes). Next, there is an optional step to convert the lyrics into meaning categories (e.g., mood labels). This requires a separate set of lyrics that have been tagged with the corresponding labels. Annotations provided by UCSD's Computer Audition Laboratory[2] are used for this purpose, specifically the CAL500 data set [9]. The mapping process uses text categorization with word features and also semantic categories in the form of WordNet ancestors [7]. Prior to categorization, both the CAL500 training data and Usenet testing data are preprocessed to isolate punctuation. However, no stemming is done (for simplicity). The remaining steps are always done. The second-last step computes contingency tables for the co-occurrence of chords and target tokens. Then these are used in the final step to derive co-occurrence statistics, such as mutual information.

```
[C] They're gonna put me in the [F] movies
[C] They're gonna make a big star out of [G] me
We'll [C] make a film about a man that's sad
    and [F] lonely
And [G7] all I have to do is act [C] naturally
```

Figure 2: Chord annotation sample. Lyrics are from "Act Naturally" by Johnny Russell and Voni Morrison, with chord annotations for song as recorded by Buck Owens.

```
C   They're  gonna  put  me  in  the
F   movies  <endl>
C   They're  gonna  make  a  big  star  out  of
G   me  <endl>  We'll
C   make  a  film  about  a  man  that's  sad  and
F   lonely  <endl>  And
G7  all  I  have  to  do  is  act
C   naturally  <endl> <endp>
```

Figure 3: Sample chord annotations extracted from lyrics. Each chord instance in figure 2 has a separate line.

3.1 Lyric Chord Annotation Data

The most critical resource required is a large set of lyrics with chord annotation. These annotations are often specified in-line with lyrics using brackets to indicate when a new chord occurs. Figure 2 shows an example. The Usenet group *alt.guitar.tab* is used to obtain the data. This is done by issuing a query for "CRD", which is the name for this type of chord annotation. The result is 8,000+ hits, each of which is then downloaded. The chord annotation data is used as is (e.g., without normalization into key of C).

After the chord-annotated lyrics are downloaded, post-processing is needed to ensure that user commentary and other additional material are not included. This is based on a series of regular expressions. The lyrics are all converted into a format more amenable for computing the co-occurrence statistics, namely a tab-separated format with the current chord name along with words from the lyrics for which the chord applies. There will be a separate line for each chord change in the song. Figure 3 illustrates this format. This shows that special tokens are also included to indicate the end of the line and paragraph (i.e., verse).

3.2 Optional Mapping via Lyric Classifier

Rather than just using the words from lyrics as the meaning content, it is often better to use terms typically associated with songs and musical phrases. This would eliminate idiosyncratic associations between chords and words that just happen to occur in lyrics for certain types of songs. More importantly, it allows for better integration with music recommendation systems, such as by using the music labels employed by the latter.

A separate dataset of lyrics is used for lyric classification. Although the overall process is unsupervised, it incorporates a mapping from words to categories based on supervised lyric classification. The source of the tagged data

[1] See http://www.last.fm.

[2] See http://cosmal.ucsd.edu/cal/projects/AnnRet.

CAL500 Emotion Categories

Label	f	Label	f
Angry-Aggressive	31	Laid-back-Mellow	7
Arousing-Awakening	77	Light-Playful	1
Bizarre-Weird	7	Loving-Romantic	1
Calming-Soothing	91	Pleasant-Comfortable	3
Carefree-Lighthearted	28	Positive-Optimistic	0
Cheerful-Festive	9	Powerful-Strong	3
Emotional-Passionate	23	Sad	3
Exciting-Thrilling	2	Tender-Soft	2
Happy	6		

Table 1: Frequency of categories from CAL500 used during classification. This reflects the frequency (*f*) of the categories for which lyrics were obtained. Only one category was applied per song, using first tag above a given threshold.

```
movie#1, film#1, picture#6, moving picture#1, ...
  => show#3
       => social event#1
            => event#1
                 => ...
       => product#2, production#3
            => creation#2
                 => artifact#1, artefact#1
                      => whole#2, unit#6
                           => ...
```

Figure 4: WordNet hypernyms for 'movie'. This is based on version 2.1 of WordNet. The first entry omits four variants in the synonyms set (e.g., flick#3), and each branch omits three levels of ancestors (e.g., entity#1).

is CAL500 [9], which uses 135 distinct categories. Several of these are too specialized to be suitable for music categorization based on general meaning, such as those related to specific instruments or vocal characterization. Others are usage related and highly subjective (e.g., music for driving). Therefore, the categorization is based only on the emotion categories. Table 1 shows the categories labels used here. Although relatively small, CAL500 has the advantage of being much more reliable than tags derived from social media like Last.fm. For instance, CAL500 uses a voting scheme to filter tags with little agreement among the annotators.

Out of the 500 songs annotated in CAL500, only 300 are currently used due to problems resolving the proper naming convention for artist and song in Lyric Wiki[3]. In addition, CAL500 provides multiple annotations per file, but for simplicity only a single annotation is used here. The resulting frequencies for the categories are shown in table 1.

Categorization is performed using CMU's Rainbow [4]. Features are based both on words as well as on semantic classes akin to word senses. WordNet ancestors called "hypernyms" [7] are used to implement this. See figure 4 for an example. The use of these word classes is intended to get around data sparsity issues, especially since the training set is rather small. The idiosyncratic nature of lyrics compared to other types of text collections makes this problem more prominent.

Contingency Table Cells

X \ Y	+	-
+	XY	$X\neg Y$
-	$\neg XY$	$\neg X\neg Y$

G versus 'film'

	+	-
+	1	2,213
-	0	17,522

Table 2: Contingency tables. The left shows the general case, and the right shows the data for chord G and 'film'.

As no part of speech tagging is applied as well as no sense tagging, the hypernyms are retrieved for all parts of speech and all senses. For example, for 'film', seven distinct senses would be used: five for the noun and two for the verb. In all, 43 distinct tokens would be introduced. Naturally, this introduces much noise, so TF/IDF filtering is used to select those hypernyms that tend to only occur with specific categories. (See [3] for other work using hypernyms in text categorization.)

Each line of the extracted chord annotations file (e.g., figure 3) is categorized as a mini-document, and the highest-ranking category label is used or N/A if none applicable. To allow for more context, all of the words from the verse for the line are included in the mini-document. The final result is a revised chord annotation file with one chord name and one category per line (e.g., figure 3 modified to have Light-Playful throughout on the right-hand side).

3.3 Chord Sequence Token Co-occurrence

Given the chord annotations involving either words or meaning categories, the next stage is to compute the co-occurrence statistics. This first tabulates the contingency table entry for each pair of chord and target token, as illustrated in table 2. (Alternatively, chord sequences can be of length four, as discussed later. These are tabulated using a sliding window over the chord annotations, as in n-gram analysis.) This table shows that the chord G co-occurred with the word 'film' once, out of the 2,213 instances for G. The word itself only had one occurrence, and there were 17,522 instances where neither occurred. Next, the *average mutual information* co-occurrence metric is derived as follows:

$$\sum_{x}\sum_{y} P(X=x, Y=y) \times log_2 \frac{P(X=x, Y=y)}{P(X=x) \times P(Y=y)}$$

4. ANALYSIS

At the very least, the system should be able to capture broad generalizations regarding chords. For example, in Western music, major chords are typically considered bright and happy, whereas the minor chords are typically considered somber and sad.[4] Table 3 suggests that the chord meaning induction process indeed does capture this generalization. By examining the frequency of the pairs, it can be seen that most cases shown fall under the major-as-happy versus minor-as-sad dichotomy. There are a few low-frequency exceptions, presumably since songs that are sad do not just restrict themselves to minor chords, as that might be too dissonant.

The exceptions shown in the table might also be due to the conventions of chord theory. In particular, chord progressions for a specific key should just contain chords based

[3] See http://lyrics.wikia.com.

[4] Strictly speaking, it is the difference in major versus minor key, but there is a close relation between keys and chords.[8]

avMI	Chord	Word	XY	$X\neg Y$	$\neg XY$
.00034	C	happy	7	1,923	13
.00005	G	happy	4	2,210	16
.00030	Dm	happy	3	341	17
.00008	Em	happy	2	548	18
.00176	F	bright	10	971	3
.00018	Am	bright	3	962	10
.00071	Bm	sad	3	197	4
.00032	Bb	sad	2	325	5
.00039	Em	sad	3	1,097	6
.00542	Dm	sorrow	2	342	5
.00068	C	sorrow	2	1,928	5

Table 3: Sample major versus minor chord associations. Within each group, the entries are sorted by joint frequency (XY). The $\neg X\neg Y$ frequency is omitted (around 17,500), along with a few singleton occurrences.

avMI	Chord Sequence	Category	XY	$X\neg Y$	$\neg XY$
.0027	$D7, D7, D7, D7$	Bizarre	30	36	1,358
.0037	$Em, G, G6, Em$	Carefree	18	6	594
.0032	$D, A, A, C^\sharp min$	Carefree	14	2	598
.0032	$C^\sharp min, D, A, A$	Carefree	14	2	598
.0032	$A, C^\sharp min, D, A$	Carefree	14	2	598
.0032	$A, A, C^\sharp min, D$	Carefree	14	2	598
.0012	$D7, G, C, G$	Bizarre	14	17	1,374
.0018	$C, D7, G, C$	Bizarre	14	19	1,374
.0022	D, A, A, D	Powerful	13	8	667
.0014	C, D, C, D	Happy	13	39	502

Table 4: Most frequent chord sequence associations. The entries are sorted by joint frequency (XY), and the $\neg X\neg Y$ frequency is omitted (around 18,700). The category names are shortened from table 1.

on the following formula, given the notes from the corresponding major scale:[8]

$$Maj(or), Min(or), Min, Maj, Maj, Min, Diminished$$

Therefore, for the key of C, proper chord sequences only contain the following chords:

$$C, Dm, Em, F, G, Am, Bm, Cdim$$

Likewise, the following are for the key of G:

$$G, Am, Bm, C, D, Em, Fm, F^\sharp dim$$

For example, both Dm and Em are among the preferred chords for the key of C major (hence reasonable for 'happy').

Of course, individual chords are limited in the meaning they can convey, given that there are relatively few that are used in practice, compared to the thousands of playable chords that are possible. For example, only 60 chords account for 90% of the occurrences in the sample from Usenet (from a total about 400 distinct chords). Therefore, the ultimate test is on how well chord sequences are being treated.

For simplicity, chord sequences were limited to length four. This was chosen given the correspondence to the number of quarter-note beats in a common time measure (i.e., 4/4 time). Over 4,000 distinct 4-chord sequences were found. As 2,500 of these account for 90% of the occurrences, there is much wider variety of usage than for individual chords.

Running the co-occurrence analysis over words runs into data sparsity issues, so instead results are shown over the mood categories inferred from the CAL500 tagged data. Table 4 shows the top sequences for which a semantic label has been inferred by the classifier (i.e., without guessing based on prior probability). For the most part, the meaning assignment seems reasonable, adding more support that the process described here can capture the meaning associated with chord sequences.

5. CONCLUSION

This paper has presented preliminary research illustrating that it is feasible to learn the meaning of chord sequences from lyrics annotated with chords. Thus, a large, untapped resource can now be exploited for use in music recommendation systems. An immediate area for future work is the incorporation of objective measures for evaluation, which is complicated given that the interpretation of chord sequences can be highly subjective. Future work will also look into additional aspects of music as features for modeling meaning (e.g., tempo and note sequences). Lastly, as this approach could be used to suggest chord sequences that convey moods suitable for a particular set of lyrics, work will investigate its use as a songwriting aid.

6. ACKNOWLEDGMENTS

Dan Ponsford provided valuable feedback on the overall process, and Per Egil Kummervold offered useful suggestions. Douglas Turnbull granted access to CAL500, and Cory McKay provided scripts for downloading lyrics.

7. REFERENCES

[1] P. Fung and K. W. Church. K-vec: A new approach for aligning parallel texts. In *Proc. COLING*, 1994.

[2] X. Hu, J. S. Downie, and A. F. Ehman. Lyric text mining in music mood classification. In *Proc. ISMIR*, pages 411–6, 2009.

[3] T. Mansuy and R. Hilderman. Evaluating WordNet features in text classification models. In *Proc. FLAIRS*, 2006.

[4] A. K. McCallum. Bow: A toolkit for statistical language modeling, text retrieval, classification and clustering. www.cs.cmu.edu/∼mccallum/bow, 1996.

[5] C. McKay et al. Evaluating the genre classification performance of lyrical features relative to audio, symbolic and cultural features. In *Proc. ISMIR*, 2010.

[6] I. D. Melamed. Models of translational equivalence among words. *Computational Linguistics*, 26(2):221–49, 2000.

[7] G. Miller. Special issue on WordNet. *International Journal of Lexicography*, 3(4), 1990.

[8] C. Schmidt-Jones and R. Jones, editors. *Understanding Basic Music Theory.* Connexions, 2007. http://cnx.org/content/col10363/latest.

[9] D. Turnbull et al. Semantic annotation and retrieval of music and sound effects. *IEEE TASLP*, 16 (2), 2008.

[10] B. Whitman and D. Ellis. Automatic record reviews. In *Proc. ISMIR*, 2004.