Predicting Chords In Jazz*

Belinda Thom and Roger Dannenberg
School of Computer Science
Carnegie Mellon University
Pittsburgh, PA 15213 USA
bthom@cs.cmu.edu, rbd@cs.cmu.edu

ABSTRACT. Instance-based machine learning techniques have been applied to predict the next chord in a piece of jazz music. This study compares off-line (trained on a static database of songs) and on-line (trained on-the-fly) prediction techniques. A hybrid model combines both on-line and off-line models. On average, chords in a jazz piece are predictable 53 ± 17% of the time using the hybrid model. Successful predictions are defined as exact matches of chord root and type.

Introduction. Anticipating chord changes is a fundamental music understanding task. Whether one is listening, improvising, or composing, chord prediction plays an important role. In computer music systems, chord prediction can be used to assist in generating melodic accompaniments with voice leading, bass figures with a sense of direction, and drum figures that reflect chordal structures. Chords that do not fit predictions convey surprise or tension, which are interesting parameters for machine composition systems.

How well can one anticipate the next chord in a piece of jazz music? A sampling of seasoned musicians, at best, gives a qualitative answer. Predicting chords in a piece of jazz music is a time series prediction problem [Weigend and Gershenfeld, 1994]. Success requires that the underlying process adheres to a predictable behavior. Consequently, prediction is limited to the extent to which the forces driving a process obey some structure and by how much underlying state is available in the training data.

Representation. Rather than using knowledge-based methods, harmonic structure is inferred directly from a database of jazz chord progressions. The machine learning task includes: selecting and encoding the feature set; defining the target concept; determining a performance measure that reflects the degree to which the target concept is learned; compiling representative train and test sets; and designing a learning algorithm [Mitchell, 1996].

Jazz chord prediction limits our choice of learning algorithms and adds a few twists to the task of selecting data and measuring performance. For example, chords are ordinal. The lack of a continuous similarity measure is a serious handicap. This lack precludes the use of continuous learning algorithms (e.g. nearest neighbor, function approximation) and hinders discrete learning algorithms because performance feedback is binary (exact match or terrible error). In addition, incremental learning is absolutely necessary given that the concept driving a progression changes within a song and among different songs. Incremental learning must be very fast since songs are short and the computer should not miss too many chords in a row. Ideally, performance feedback to the learning algorithm would include: temporality (how are the errors spaced over time?), context (how bad are the errors musically?), and degree (how far off is each chord?). This type of feedback requires the same level of musical understanding we are trying to learn! Selecting train and test sets are also problematic. Usually, data is drawn at random from the universe. We are interested in predicting harmonic structure on a per-song basis. Within a song, harmonic structure is highly localized. Predictive performance observed for one song is not necessarily a good indicator of another song's performance. It is also unclear how to normalize songs in the training set so the per-song universe is represented optimally. All of these issues need to be researched further.

The problem is restricted to predicting the next unique chord given a chord progression history. Fifty random songs selected from a jazz fake book comprise the test and train sets. All songs are transposed to the key of C; superfluous repeats are removed. Chord types are mapped to a simplified set: Major, Minor, Dominant, Diminished, Augmented, Other. Features sets are a function of history window: n most-recent-chords, $0 \leq n \leq 3$. Several encodings are considered:

- *Absolute (Abs)*: chord histories are referenced to absolute pitch (e.g. {c-Maj, f-Maj, g-7});

*Published as: Belinda Thom and Roger B. Dannenberg, "Predicting Chords in Jazz," in *Proceedings of the 1995 International Computer Music Conference,* Banff, Alberta, Canada, September 2005. International Computer Music Association, 1995. pp. 237-238.

- *Interval-A (Int A)*: chord histories are referenced to previous pitch (e.g. {IV-Maj, IV-Maj, I-7});
- *Interval-B (Int B)*: same as Int-A, except the earliest feature is chord type (e.g. {Maj, IV-Maj, I-7}).

Performance is quantified by number of exact matches predicted for an entire song. Cross-validation is used to obtain a robust performance metric. Forty-nine songs comprise the training set; the remaining song is held out as the test set. By repeating this train-and-test procedure on each of fifty songs, an average is obtained that quantifies the performance one would expect on a new song.

Our learning algorithms are based on n-gram models [Bell, *et al.* 1990]. Prediction is determined by maximizing the likelihood of a chord progression with respect to a database. The ultimate goal is to create an expert jazz predictor that adapts to the vagaries of a song's local structure. This goal is accomplished as follows. Initially nothing is known about a particular song; off-line data is consulted. As on-line data appears, the off-line data set incorporates this data, favoring it heavily, 50:1. The success of this hybrid scheme is intimately tied to: the underlying structure captured by small, finite-context models; the degree to which a song is covered by the off-line data; and the degree to which a song's on-line training examples cover itself.

Results. Experiments were run to answer the following questions:

1. *How well does our limited representation capture harmonic structure?* To answer this question, an off-line database is trained using only the test song. As one would expect, performance improves as n increases. In this experiment, performance is directly tied to the entropy a song has given a particular feature set and encoding. Most songs perform best with the Abs encoding, although for some Int A is optimal. It is surprising how well our limited representation performs: on average, chords are uniquely determined $92 \pm 4\%$ of the time when a song is its own training set. This experiment is the best possible case, providing an upper bound for comparing other experiments.

2. *How much harmonic similarity exists among jazz songs?* For each of the fifty songs, an off-line database is trained with all but the test song; performance of the test song is observed using this database. This jazz database is surprisingly ineffective. Optimal history size and encoding vary wildly from song to song. Often, significant portions of a song are either incorrectly covered or not covered at all. Given a song's optimal representation (either Abs, Int A, or Int B, and a specific n) prediction is correct $42 \pm 17\%$ of the time (this value is misleading because it avoids the adaptive issue of determining optimal representation on-the-fly). The poor results of this experiment are attributed to over-fitting and lack of on-line adaptation.

3. *How well does our representation capture local structure?* Performance of each song is observed using an empty database trained on-the-fly. Structure is quickly inferred. Once again, performance improves as n increases and the Abs encoding is often optimal. *In jazz, incremental learning alone outperforms a fifty song knowledge base*; on average, predictions are correct $49\% \pm 15\%$ of the time.

4. *What advantage is obtained by combining on-line and off-line data?* The hybrid model is used for these experiments. For each test song, the initial database contains the other forty-nine songs. On-line data is incorporated with a 50:1 weighting. Although the Abs encoding and larger n tend to improve performance, per-song variance warrants investigating how well adaptive techniques can be applied to determining optimal representations on-the-fly. Hybrid models do improve prediction. On average, given the optimal representation, prediction is correct $53 \pm 17\%$.

Conclusions. The most useful result of this study is what it says about the similarities that arise in jazz chord progressions. Although patterns are shared among songs, local data seems to provide much more reliable predictive information. This study is also useful because it bounds the level of performance one can expect when restricted by the exact match constraint. This study also indicates that computer jazz chord prediction succeeds because songs have highly repetitive, localized structure.

Ultimately, the essence of jazz performance is interactive and improvisational. Improving the computer's ability to anticipate chords and understand chord structure is critical for its active participation in this art form.

References.

Bell, T. C., J. G. Cleary, and I. H. Witten. 1990. *Text Compression.* Prentice Hall Advanced Reference Series, New Jersey: Prentice Hall.

Mitchell, T. 1996 (to appear). *The Essentials of Machine Learning.*

Weigend A. and Gershenfeld. N. 1994. *Time Series Prediction: Forecasting the Future and Understanding the Past.* Massachusettes: Addison-Wesley.