A MEMORY-BASED APPROACH TO METER INDUCTION

Menno van Zaanen [1,2], Rens Bod [2,4]& Henkjan Honing [2,3]
[1] ILK, University of Tilburg, Netherlands
[2] ILLC, University of Amsterdam, Netherlands
[3] Music Department, University of Amsterdam, Netherlands
[4] University of Leeds, UK

ABSTRACT

Meter induction has been an important topic in the computational modeling of music cognition for quite some time now. In this paper, an attempt is made to model how listeners arrive at a metrical interpretation of a fragment of music. A number of existing models are based on the Gestalt principles of perception, 'simplicity' or ease of encoding being a key aspect. An alternative to this approach are models based on the notion of 'likelihood', so-called memory-based models. We adapt and evaluate a number of memory-based models for parsing metrical structure. More specifically, we will use the models covered by the Data-Oriented Parsing (DOP) framework. This framework defines a large class of probabilistic grammars that take sub-trees from an annotated corpus to form a general Probabilistic Tree Grammar. The models are tested on the National Anthems collection, yielding encouraging results.

1. INTRODUCTION

Even though the computational modeling of beat and meter induction has been researched for some time now (Desain & Honing, 1994), the human assignment of metrical information still outperforms existing computational models and systems. Humans are not only very precise in finding structural metrical information, they can also do it quickly and are very flexible, i.e. they can easily adapt to meter changes. There is a considerable amount of literature on modeling the phenomenon of meter induction, using a large variety of computational paradigms (cf. Desain & Honing, 1999). One class of models is based on the Gestalt principles of perception, 'simplicity' or ease of encoding being a key aspect. An alternative approach, called memory-based, is based on the notion of 'likelihood'. Here, models try to explain structural interpretations in terms of the most probable encoding. The probabilities are extracted from previously seen examples. Instead of generating the metrical structure using a simple model, previously encountered structures drive the analysis of new data.

In this paper, we explore a number of memory-based approaches to meter induction concentrating on models that fit the Data-Oriented Parsing (DOP) framework (see Bod, Scha & Sima'an, 2003 for an overview). The models are tested on the National Anthem Collection (Desain & Honing, 1999). We will show that (fragments of) previously seen examples of metrical information can be used to assign structure to an unseen piece of music (see Figure 1).

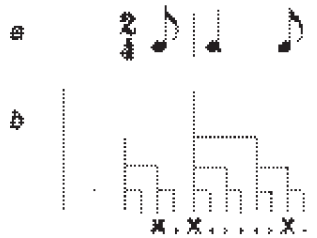

Figure 1: A rhythmic example in common music notation (**a**), and its representation on a time grid (with x for an onset, dot for a rest, one grid-point is a 16th note), and with a metrical tree above it (**b**).

The results obtained indicate that memory-based approaches can be considered as a viable alternative to existing models of meter induction.

The next section explains several memory-based approaches to the structuring of data, concentrating on the family of DOP-models. Section 3 contains a description of how these models are applied and evaluated, together with the actual results.

2. MEMORY-BASED APPROACH

The DOP-framework (Bod, 1998; Bod, Scha & Sima'an, 2003) defines a large class of probabilistic grammars by taking sub-trees from an annotated corpus to form a general Probabilistic Tree Grammar. Sub-trees are formed by 'cutting' the original tree structure on all possible internal nodes. Cutting the trees from the training data results in many different sub-trees, where each may possibly occur more than once. The extracted sub-trees have open nodes (i.e. non-terminals rather than terminals), that generalize over the original tree structure by under-specifying parts of the complete tree, effectively creating the generative power of the approach.

By limiting the sub-trees in various ways, several specific probabilistic grammars can be simulated (e.g., by limiting the sub-trees to depth 1, a probabilistic context-free grammar or PCFG is obtained). Thus the underlying idea of DOP is to analyze new data using sub-trees from a corpus of previously analyzed data.

ISBN 3-931852-67-9
ISSN 1617-6847

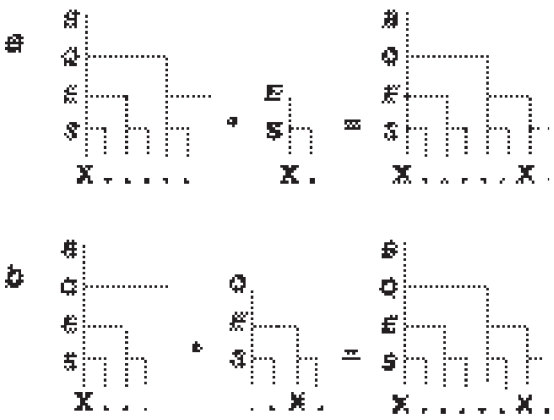

Figure 2: An example of how sub-trees can combine in a fully specified metrical interpretation. Using attached probabilities, the most probable tree structure can be chosen; **a** and **b** show two different ways of arriving to the same tree.

A node-substitution operation is used to combine two or more sub-trees into a new tree structure (see Figure 2). Node-substitution identifies the leftmost non-terminal frontier node of one sub-tree with the root node of a second sub-tree (i.e., the second sub-tree is *substituted* on the leftmost non-terminal frontier node of the first sub-tree). A sequence of sub-trees that can be successfully combined is called a *derivation*, while the structure resulting from a derivation is called a *parse tree*. The most probable parse tree for an input is compositionally computed from the probabilities of the sub-trees. The probability of a sub-tree t, $P(t)$, is computed as the number of occurrences of t, $|t|$, divided by the total number of occurrences of treebank-sub-trees that have the same root label as t. Let $r(t)$ return the root label of t. Then we may write:

$$P(t) = \frac{|t|}{\sum_{t' : r(t')=r(t)} |t'|}$$

The probability of a derivation $t_1 \circ ... \circ t_n$ is computed by the product of the probabilities of its sub-trees t_i:

$$P(t_1 \circ ... \circ t_n) = \prod_i P(t_i)$$

There may be different derivations that generate the same parse tree (see Figure 2a/b). The probability of a parse tree T is the sum of the probabilities of its distinct derivations. Let t_{id} be the i-th sub-tree in the derivation d that produces tree T, then the probability of T is given by

$$P(T) = \sum_d \prod_i P(t_{id})$$

Thus the DOP method considers counts of sub-trees of a wide range of sizes in computing the probability of a parse tree: everything from counts of single-level rules to counts of entire trees. This means that the method is sensitive to the frequency of large sub-trees while taking into account the smoothing effects of counts of small sub-trees.

Standard best first parsing algorithms can be applied to computing the most probable parse tree for an input in DOP (for details see Bod, Scha & Sima'an, 2003).

3. RESULTS

This section presents an evaluation of the memory-based approaches to meter induction. First, the data sets are described, followed by an explanation of the settings and the evaluation approach taken. Finally, some quantitative results are discussed.

3.1. Data Sets

The memory-based models are tested on the National Anthems Collection (Desain & Honing, 1999), or *Anthems* for short: a metrically annotated corpus (see http://www.hum.uva.nl/mmm under heading 'Data Archives'). This collection contains 105 songs (see Table 1). The collection is pre-processed to obtain a labeled metrical tree description of each Anthem (cf. Figure 1). Figure 3a depicts the structure of an Anthem (i.e. an example of a 2/4 interpretation). The highest level delimits the piece (P), the next highest level denotes bar information (B). The levels below that contain duration information. Depending on the meter, it contains half (H), quarter (Q), eighth (E), and sixteenth (S) notes. All anthems are notated on a sixteenth note time grid.

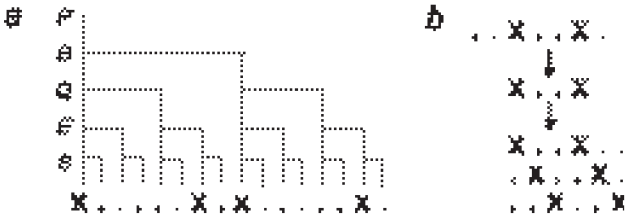

Figure 3a: A rhythmic example of two bars making up a song. The levels are labeled P(iece), B(ar), Q(uarter-note), E(ighth-note) and S(ixteenth) note respectively. **b**: An example of an input string (in time-grid representation) and its alternative versions filled-out to make-up a pattern of n-grid units long (here n=6) to allow for all possible upbeat interpretations.

Using entire songs for testing and training raises two issues. Firstly, since humans can assign metric information already after analyzing a very limited amount of music (Desain & Honing, 1999), it seems only fair to allow a model of human musical perception access only to a fragment of a song, too. Secondly, the memory-based models that are evaluated in this article are computationally limited with regard to the length of the test and training data. Large training trees result in too many possible sub-trees (the number of sub-trees grows exponentially with the size of the original tree). Similarly, structuring relatively long sequences is currently outside the practical boundaries in memory space and computational power.

Bearing this in mind, we choose to select for this study the first 48 grid-points (denoting 16th notes) of each song for testing and training. Table 1 contains a list of the meters that are found in the collection, together with their distribution. In all of these meters, 48 sixteenth notes fit within full bars perfectly. For example, with a 3/4 meter, a bar contains 12 sixteenth notes, so 48 grid-points delimit 4 bars.

ISBN 3-931852-67-9
ISSN 1617-6847

Meter	N	Proportion
4/4	77	0.73
2/4	11	0.11
3/4	10	0.10
2/2	6	0.06
6/4	1	0.01

Table 1: Distribution of meters in Anthems corpus (N=105)

The first 48 grid-points of a test song may start with a rest, signifying an upbeat. Since the system should be able to recognize upbeats, all grid-points with rests at the beginning and end of these grid-points are removed. This yields a shorter string of grid-points starting and ending in a note onset. Next, the set of 48 grid-point songs containing all possible upbeats of the reduced excerpt is generated by sliding the excerpt over an 48 grid window. Rests are padded to the left and right to fill up the missing parts. In other words, rests are positioned in front of the excerpt and at the end, filling an 48 grid song (see Figure 3b). Note that this is not the only method of generating possible upbeats. However, the main advantage of this approach is that the number of rests in the pieces stays the same. Since most training songs have upbeats and the upbeats are a significant piece of the songs, the system has a preference for songs containing rests. By keeping the number of rests in all possible songs the same, this unwanted preference is diminished (NB. In future research, we will investigate alternative methods of upbeat generation). It must be stressed that this is not a property of the approach per se, but a combination of the selection and annotation of the training data, and the small size of the excerpts.

3.2. Evaluation

We divided the Anthem Collection into 10 different training/test set splits, where 10% of the songs were used as test data each time, and 90% as training data. The test data, consisting of pieces with upbeats of different lengths, are handed to the system. Based on the sub-trees extracted from the training data, the test data songs are parsed and ranked by the *dopdis* parser, implementing the DOP framework (Sima'an, 1999). The pieces with the highest probabilities are selected (in effect, selecting the most probable upbeat) and are returned as output of the system. This is then evaluated against the original or *gold standard* structure. This comparison can be done in many different ways. In this paper we test whether both the bar-length and upbeat-length are identical. But note that the memory-based systems find complete metrical tree structures (not just phase and duration of a beat) describing the complete metrical structure. For now, we concentrate on the beat-level only. Future work will further investigate generated sub-bar information.

3.3. Quantitative Results

The results of applying the *dopdis* parser (Sima'an, 1999) to the Anthem Collection can be found in Table 2. The first column denotes the maximum depth of the training set sub-trees used. When depth 1 is used, the system is equivalent to using a probabilistic context-free grammar (PCFG). The second column, 'Upbeat' denotes the percentage of correctly found upbeats. '1st Bar' gives the percentage of correct bar-length of the first bar. '2nd Bar' gives the percentage of correct bar-length of the second bar, while 'Any Bar' gives the percentage of Anthems containing a correct bar-length. The figures between brackets are the standard deviation rates.

Depth	Upbeat	1st Bar	2nd Bar	Any Bar
1	36 (6.36)	5 (2.24)	1 (1.00)	6 (2.67)
2	49 (4.58)	9 (3.48)	22 (3.59)	37 (3.96)
3	57 (3.96)	1 (3.14)	39(4.82)	50 (2.98)

Table 2: Results of *dopdis* parsing on Anthems corpus.As can be concluded from the results (See Table 2), the dopdis parser is not as successful as, for instance the family of rule-based systems described in Desain & Honing (1999) in finding correct upbeats; the latter finds up to 60% correct beats, as the dopdis parser finds up to 50% correct.

A possible reason for this result is that any combination of note onsets and rests can be parsed by the DOP framework. The selection of the song with the correct upbeat depends entirely on the statistics of previously structured data, and therefore rule-based methods (as described in Desain & Honing, 1999) are likely to do better. However, as one can see, as the maximum sub-tree depth increases (See Table 2, first column), more structural and statistical information is gathered, clearly leading to better results.

In future research, we hope to increase the performance of our models by further enlarging structural context (such as allowing larger sub-trees) and by using more sophisticated probabilistic training algorithms (such as expectation-maximization and maximum entropy; see Bod, Scha & Sima'an, 2003). Furthermore, a more thorough investigation into the assignment and generation of possible upbeats is asked for. The current method is rather ad hoc, but other approaches will be considered.

4. CONCLUSION

In this article, we presented a memory-based approach to meter induction. The family of systems that employ this approach use knowledge extracted from previous experience to analyze new and possibly unseen instances. We concentrated on the Data-Oriented Parsing (DOP) family of models. The DOP systems can be applied to metrically annotated corpora of music, although some problems arise. The main problem of the system turned out to be determining of the upbeat (or phase) in a straightforward manner. The approach taken here is to analyze a set of possible upbeats and let the system select the most probable upbeat based on the probability of the structure found. The results show that when more structural and probabilistic information is used (i.e. when a larger maximum tree depth is used), the results increase significantly. A potential advantage of the DOP approach is that it can in principle take into account changes in meter, which the rule-based approaches mentioned do not address. Experiments with irregular meter will, however, have to await further experimentation. In future research we will, next to the issues mentioned, apply these methods to larger corpora such as the Essen Folksong Collection (Schaffrath, 1995).

ISBN 3-931852-67-9
ISSN 1617-6847

5. ACKNOWLEDGMENTS

We would like to thank Khalil Sima'an for making his *dopdis* parser (owned by the Netherlands Organization of Scientific Research [NWO]) available and for his extensive help in configuring it.

6. REFERENCES

1. Bod, R.(1998) *Beyond Grammar*. Stanford, CA: CSLI Publications.

2. Bod, R., Scha, R. & Sima'an, K (eds.)(2003). *Data-Oriented Parsing*, University of Chicago Press.

3. Desain, P., & Honing, H. (1994). Foot-Tapping: a brief introduction to beat induction. In *Proceedings of the 1994 International Computer Music Conference*. 78-79. San Francisco: International Computer Music Association.

4. Desain, P. & Honing, H. (1999) Computational Models of Beat Induction: The Rule-Based Approach. *Journal of New Music Research* **28**(1):29-42.

5. Schaffrath, H. (1995) *The Essen Folksong Collection in the Humdrum Kern Format*. D. Huron (ed.). Menlo Park, CA: Center for Computer Assisted Research in the Humanities.

6. Sima'an, K. (1999) *Learning Efficient Disambiguation*. PhD thesis. University of Amsterdam / University of Utrecht, The Netherlands.

ISBN 3-931852-67-9
ISSN 1617-6847