

N-GRAM PATTERN MATCHING AND DYNAMIC PROGRAMMING FOR SYMBOLIC MELODY SEARCH

Alexandra L. Uitdenbogerd

RMIT University
Melbourne, Victoria
Australia

ABSTRACT

For this submission to MIREX 2007, we again provide a simple base-line for comparison against other algorithms for the task of symbolic melody matching against both monophonic and polyphonic collections of music. This year, we have included the n-gram-based matching technique that is implemented by building an n-gram index of the query, which is then used to search through each melody or track within the collection. In addition we provide the two dynamic programming algorithms submitted to MIREX 2006 for the same tasks. All three algorithms were statistically indistinguishable from the other algorithms submitted for the task, and were two orders of magnitude faster.

1 INTRODUCTION

In the interests of providing continuity so that algorithms can be compared across years, we again submit algorithms to be used as a baseline for the evaluation of symbolic music matching.

In our prior work on the topic of symbolic music matching [1, 2, 3] we found that the use of n-grams of length four to seven were about as effective as a dynamic programming-based matching technique (local alignment) for finding relevant melodies in a collection of polyphonic symbolically stored music. The melodies were matched using an intermediate form consisting of strings that encoded the interval between adjacent notes, with a maximum interval of one octave. Intervals larger than an octave were mapped to the harmonically equivalent interval within an octave.

The n-gram technique can be implemented in a variety of ways. For MIREX 2005, an inverted index was used, which is theoretically the most efficient approach for a large number of queries. The index building cost, however, can be quite large. For this submission, we have chosen to submit an implementation that has low index-building costs but still has much faster search than dynamic programming-based matching.

2 TECHNIQUES

All three techniques submitted to MIREX 2007 make use of the three-phase music matching model, which consists of melody extraction, melody standardisation, followed by melody matching [3, 4]. The melody extraction phase, includes notes that are the highest pitch at each instant. Melody standardisation converts the note sequence to a sequence of intervals that have a maximum size of one octave, with all intervals exceeding that interval being mapped to a harmonically similar interval within an octave. For example, the interval from D to G an octave and a fourth (perfect 11th, or 17 semitones) would be mapped to the interval of a perfect fourth (5 semitones). Our experiments on symbolic polyphonic collections of approximately then thousand pieces showed little difference in precision when matching using an exact interval representation and the simplified representation described above, that we call “directed modulo-12” [3].

The third and final stage of the process consists of matching the standardised query melody to each of the standardised melodies of the collection. The three matching techniques are described below.

2.1 Matching Techniques

The first matching technique, which we have named *Start-Match Alignment*, initialises and fills the array in the manner of global alignment, but, in the manner of local alignment, returns the highest score within the matrix. The equation used to calculate each cell’s value is the same as for global alignment.

$$a[i, j] = \max \begin{cases} a[i-1, j] + d & i \geq 1 \\ a[i, j-1] + d & j \geq 1 \\ a[i-1, j-1] + e & p(i) = t(j) \text{ and } i, j \geq 1 \\ a[i-1, j-1] + m & p(i) \neq t(j) \text{ and } i, j \geq 1 \\ 0 & i, j = 0 \end{cases} \quad (1)$$

where d is the cost of an insert or delete, e is the value of an exact match, m is the cost of a mismatch, i and j are non-negative integers, $p(i)$ represents the i th symbol in the “pattern” or query, and $t(j)$ represents the j th symbol in the “text”, or potential answer string. The weights we

used were 1 for a match, -1 for a mismatch, and -2 for an insert or delete (indel).

In addition, we have submitted a simple local alignment algorithm. The equation is similar to the above, including the specific weights used. The only difference is that each cell has a minimum value of 0.

The final technique used is our distinct n-gram counting technique. We used n-grams of length 5. Each distinct n-gram that occurs in both the melody and the answer (track) is counted. In the field of text retrieval, this is equivalent to “coordinate matching”.

3 MIREX TASKS AND RESULTS

The three algorithms were applied to the MIREX 2007 Symbolic Melody Matching task. The collection to be searched was the Essen collection of 5274 folk song melodies. Six queries were selected from the set, evaluations of relevance made by six human assessors (one query per assessor), and 4 variants of each query were tested, with relevance assumed to be the same as that of the original queries.

The average precision and precision at N scores show that for queries that have a single note inserted or deleted, the start-match (MIREX Algorithm 1) and local alignment (MIREX Algorithm 3) algorithms are likely to give the best results respectively. These methods appeared to be less able to recover from an enlarged or compressed interval, however, than other algorithms submitted to the symbolic melody matching track.

The statistical significance scores published for the track show that our three algorithms were not statistically significantly different in effectiveness from the highest-scoring algorithms.

Of the algorithms submitted to the track, our three were the fastest by two orders of magnitude. We question the run-time of the n-gram queries, however, as this typically runs 5–20 times faster than the dynamic-programming-based algorithms on our machines and collections. It is possible that a decimal place is missing in the published times.

4 DISCUSSION

Given the results in the current MIREX round it would be difficult to recommend a technique other than our own for symbolic melody search. However, there are some potential weaknesses with the experimental design which probably have affected the outcome. First, the query set was a little small, albeit expanded to a set of 30 via the introduction of different types of error. Second, the relevance assessments appear to have been based on the original exact query, rather than the versions with errors. It is possible that some of the errors could have resulted in different relevance assessments by human assessors.

The Start-Match algorithm, if statistical significance is ignored, was again shown to be better than local alignment for melody matching. Precise details about how the

query was created have not been provided. If the queries were incipits, then this would explain the superior performance of Start-Match. Otherwise, there may be a human tendency to be more aware of similarity that occurs at the start of a pair of melodies than elsewhere. This concept is in line with research in learning, where the first things learnt are more remembered than subsequent knowledge (primacy).

5 CONCLUSION

The MIREX event is a worthwhile activity for the comparison of algorithms for music retrieval and well worth supporting by a larger cohort of researchers. However, for more meaningful results to be achievable, larger query sets are probably required.

The baseline algorithms that we have submitted this year are still competitive in both effectiveness and efficiency. The algorithms are well-documented in the literature, and the code for one of them is available online¹. We recommend that researchers use them as a baseline when demonstrating their retrieval effectiveness for melody search in symbolic collections.

6 REFERENCES

- [1] A. L. Uitdenbogerd. *Music Information Retrieval Technology*. PhD thesis, School of Computer Science and Information Technology, RMIT University, Melbourne, Victoria, Australia, 2002.
- [2] A. L. Uitdenbogerd, A. Chattaraj, and J. Zobel. Methodologies for evaluation of music retrieval systems. *INFORMS Journal of Computing*, 18(3), 2006. (originally presented at ISMIR 2000).
- [3] A. L. Uitdenbogerd and J. Zobel. Music ranking techniques evaluated. In M. Oudshoorn, editor, *Proc. Australasian Computer Science Conference*, Melbourne, Australia, January 2002.
- [4] A. L. Uitdenbogerd and J. Zobel. An architecture for effective music information retrieval. *J. American Society of Information Science and Technology (JASIST)*, 55(12):1053–1057, 2004.

¹ See <http://mirt.cs.rmit.edu.au/fanimae/>