SIMPLE EFFICIENT *N*-GRAM INDEXING FOR EFFECTIVE MELODY RETRIEVAL

Iman S. H. Suyoto and Alexandra L. Uitdenbogerd
School of Computer Science and Information Technology, RMIT
GPO Box 2476V, Melbourne 3001
Australia
ishs@acm.org, sandrau@rmit.edu.au

ABSTRACT

RMIT MIRT Fanimae MIREX 2005 Edition (*FM05*) is a music information retrieval tool implementing several variations of a string matching approach developed for handling a wide variety of symbolic musical data. This approach consists of three phases: melody extraction, melody standardisation, and similarity measurement. This implementation uses the directed modulo-12 standardisation technique — a simplified pitch interval representation — that was shown to work about as well as an exact interval approach. For similarity measurement we use coordinate matching on n-grams of pitch. The software builds an *n*-gram index to allow fast search on large collections. This simple technique performed quite well compared to other submitted algorithms in the 2005 MIREX symbolic melodic similarity contest.

1 INTRODUCTION

Earlier work on symbolic melody matching (Uitdenbogerd and Zobel, 2002) showed that coordinate matching on *n*-grams of pitch interval strings worked remarkably well compared to local alignment of the same strings, being statistically indistinguishable for queries against a collection of about ten thousand MIDI files. For this MIREX entry, we chose to present an efficient index-based implementation of the technique as a base-line for the event. To our pleasant surprise this technique was ranked third of the seven algorithms, and was the fastest of the top ranking techniques using MIDI source data.

FM05 is aimed towards high effectiveness and scalability to large music databases. Therefore, *FM05* uses indexing to reduce searching cost instead of exhaustive search. The system is also highly portable. As the input, *FM05* accepts melody sequences that contain both pitch and duration information. Pitch sequences are encoded in the *directed modulo-12* standardisation, and compared using coordinate matching of 5-grams (Uitdenbogerd and Zobel, 2002).

2 STANDARDISATION

Directed modulo-12 standardisation encodes the pitch intervals between adjacent melody notes as a number of semitones, with positive for up and negative for down. Intervals larger than an octave are reduced to a harmonically equivalent interval preserving the contour. For example a leap of two octaves is reduced to one octave, and a leap of a major 10th (16 semitones) is reduced to a major 3rd (4 semitones). This standardisation technique allows matching to occur regardless of the presence or absence of bar and beat information, and seems to work well despite lacking additional information.

3 SIMILARITY MEASUREMENT

FM05 implements coordinate matching, which involves counting the distinct terms that a query and potential answer have in common. In *FM05* the terms are pitch *n*-grams of length five. For example, suppose the query is "`0 2 5 0 2 5`". This is broken into two 5-grams: "`0 2 5 0 2`" and "`2 5 0 2 5`". Suppose there are three tunes in the collections:

- Tune X: "`0 2 5 0 2 -2 -1 0 2 5`". The 5-grams are: "`0 2 5 0 2`" (also in the query), "`2 5 0 2 -2`", "`5 0 2 -2 -1`", "`0 2 -2 -1 0`", "`2 -2 -1 0 2`", and "`-2 -1 0 2 5`". There is one common n-gram, leading to a score of 1.
- Tune Y: "`0 2 5 0 2 2 5 0 2 5 0 2`". The 5-grams are: "`0 2 5 0 2`" (also in the query), "`2 5 0 2 2`", "`5 0 2 2 5`", "`0 2 2 5 0`", "`2 5 0 2 5`" (also in the query), "`5 0 2 5 0`", and "`0 2 5 0 2`" (second occurrence in the tune, and also in the query). There are two common *distinct* n-grams, leading to a score of 2.
- Tune Z: "`1 2 3 4 5 6`". The 5-grams are: "`1 2 3 4 5`" and "`2 3 4 5 6`". There are no common n-grams, thus a score of 0.

Tune Y is ranked first as it has the largest number of common distinct 5-grams. X comes second, and Z third.

4 EXPERIMENT

Prior to the contest, *FM05* was tested on a collection of MIDI-formatted incipits[1]. Testing was done on a 2.80-GHz Intel Pentium 4 with 1 GB of RAM. The operating system used was Linux 2.4.22. Indexing took 35 seconds, and querying using all the incipits as queries took 3 seconds.

In the contest itself, *FM05* was ranked third of seven entries in effectiveness, with an average dynamic recall of 64.18%. There is a substantial gap between the first three entries and the remainder using this measure of effectiveness. A gap also occurs for the normalised recall at group boundaries measure. For *r*-precision our technique achieved 41.72%, the second highest score, but for non-interpolated average precision our result of 40.42% was only the fourth highest result.

Figure 1 shows the average dynamic recall versus the run-time for each algorithm. Note that some algorithms cannot really be directly compared for run-time efficiency due to the different platforms and data formats used, and that some had evaluation time included in the total run-time.

5 CONCLUSION

Coordinate matching of *n*-grams is a simple and effective approach to symbolic melody retrieval, and makes a worthy baseline for researchers to use in their experiments. In the MIREX 2005 competition the advantage of building and using an index was not clear due to the small collection size of 558 files and having only 11 queries in the query set. Indeed, the dynamic programming techniques of two submitted algorithms were faster.

The two leading techniques in the contest use slightly more complex formulations than ours. One uses an edit distance of structures; the other combines various *n*-gram lengths of strings of pitch and duration. The same is true of three techniques that were ranked lower than ours. The experimental results show that care needs to be taken when modelling melodic similarity, since incorporating richer musical information does not automatically result in greater effectiveness, something we had discovered previously (Suyoto and Uitdenbogerd, 2005).

References

Iman S. H. Suyoto and Alexandra L. Uitdenbogerd. Effectiveness of note duration information for music retrieval. In Lizhu Zhou, Beng Chin Ooi, and Xiaofeng Meng, editors, *Proceedings of International Conference on Database Systems for Advanced Applications*, pages 265–275, Beijing, China, April 2005. Springer-Verlag.

Alexandra L. Uitdenbogerd and Justin Zobel. Music ranking techniques evaluated. In *Proceedings of Australasian Computer Science Conference*, pages 275–283, Melbourne, Australia, 2002.

[1] See `http://teuge.labs.cs.uu.nl/Ruu/mirex/`

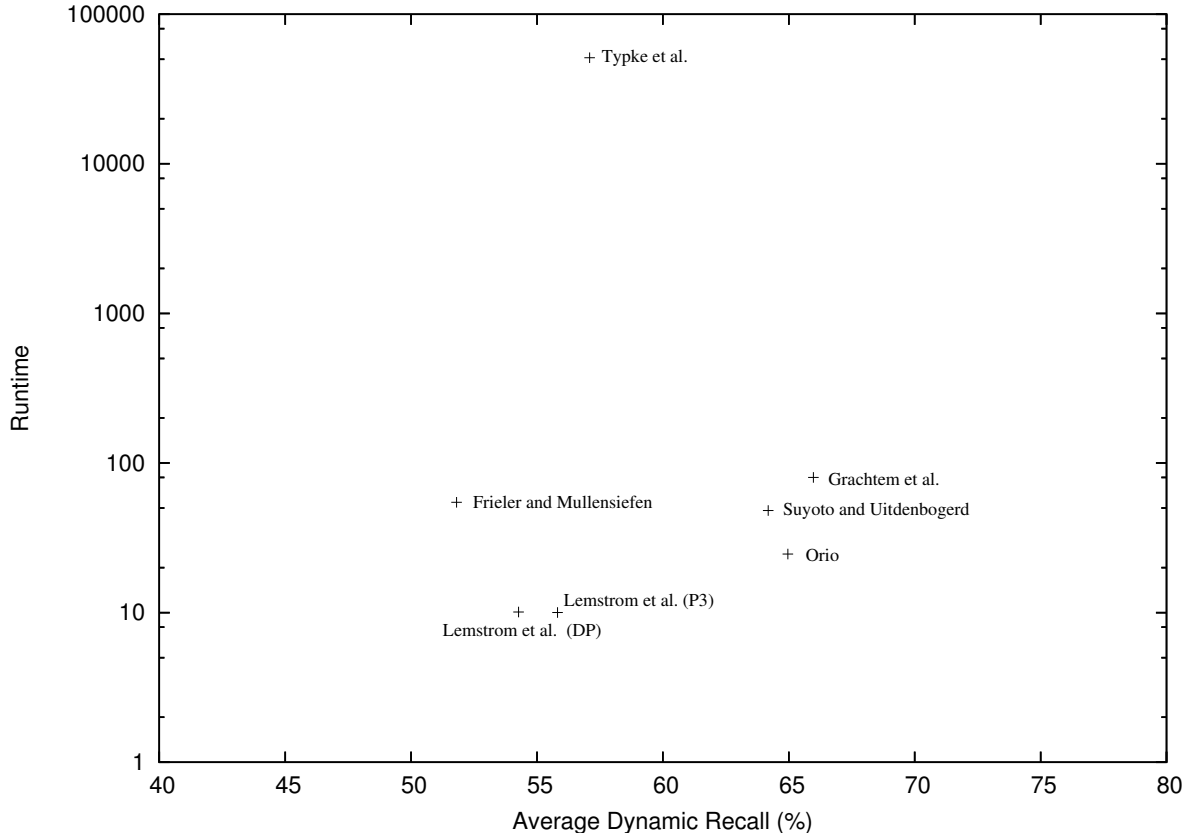

Figure 1: Average dynamic recall versus run-time for the MIREX 2005 symbolic melodic similarity contest. The best possible results would be rightmost and close to the horizontal axis.