Using Relevance Feedback to Learn Both the Distance Measure and the Query in Multimedia Databases

Chotirat Ann Ratanamahatana, Eamonn Keogh

Dept. of Computer Science & Engineering, Univ. of California, Riverside, CA 92521 USA
{ratana, eamonn}@cs.ucr.edu

Abstract. Much of the world's data is in the form of time series, and many other types of data, such as video, image, and handwriting, can easily be transformed into time series. This fact has fueled enormous interest in time series retrieval in the database and data mining community. However, much of this work's narrow focus on efficiency and scalability has come at the cost of usability and effectiveness. Here, we introduce a general framework that learns a distance measure with arbitrary constraints on the warping path of the Dynamic Time Warping calculation. We demonstrate utility of our approach on both classification and query retrieval tasks for time series and other types of multimedia data, then show that its incorporating into the relevance feedback system and query refinement can further improve the precision/recall by a wide margin.

1 Introduction

Much of the world's data is in the form of time series, leading to enormous interest in time series retrieval in the database and data mining community. However, most of previous work has utilized the Euclidean distance metric because it is very amenable to indexing [1], [2]. However, there is increasing evidence that the Euclidean metric's sensitivity to discrepancies in the time axis makes it unsuitable for most real world problems [3], [4], [5], [6]. This fact appears to be almost unnoticed because, unlike their counterparts in information retrieval, many researchers in the database/data mining community evaluate algorithms without considering precision/recall or accuracy [7]. In this work, we introduce a distance measure and show its utility with comprehensive experiments. Despite its potential weakness that it requires some training or human intervention to achieve its superior results, we can use the classic relevance feedback technique to achieve this end.

1.1 The ubiquity of time series data

In this section, we wish to expand the readers' appreciation for the ubiquity of time series data. Rather than simply list the traditional application domains, we will consider some less obvious applications that can benefit from efficient and effective retrieval.

Video retrieval: Video retrieval is one of the most important issues in multimedia database management systems. Generally, research on content-based video retrieval represents the content of the video as a set of frames, leaving out the temporal features of frames in the shot. However, for some domains, it may be fruitful to extract time series from the video, and index just the time series (with pointers back to the original video). Figure 1 shows an example of a video sequence transformed into a time series. One obvious advantage of this representation is the huge reduction in dimensionality, enhancing the ease of storage, transmission, analysis, and indexing; it is also much easier to make the time series representation invariant to distortions in the data.

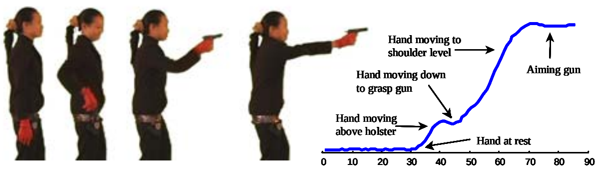

Figure 1. Stills from a video sequence; the right hand is tracked, and converted into a time series

Image retrieval: Image Retrieval has become increasingly crucial in our information-based community. For some specialized domains, it can be useful to convert images into “pseudo time series.” Consider Figure 2, where we convert an image of a leaf into a time series by measuring the local angle of its perimeter’s trace. The utility of such a transform is similar to that for video retrieval, especially in natural images, where we can more easily handle the natural variability of living creatures (scale, offset, and rotation invariance) in the time domain with dynamic time warping [8].

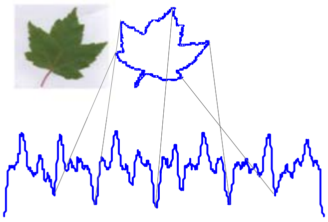

Figure 2. Many image indexing/classification/clustering can be solved more effectively and efficiently after converting the image into a "pseudo time series"

Handwriting retrieval: While the recognition of online handwriting [9] may be largely regarded as a solved problem, the problem of transcribing and indexing existing historical archives remains a challenge.

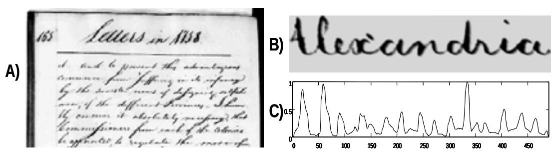

Figure 3. A) An example of handwritten text by George Washington. B) A zoom-in on "Alexandria", after slant removal. C) A variety of techniques exist to convert 2-D handwriting into a time series; in this case, the projection profile is used

The problem of indexing historical archives is difficult, because unlike the online handwriting problem, there is no pen-acceleration information. In addition, the archives may be degraded and stained. For example, Figure 3.A) shows an example of text written by George Washington, which is all but illegible to modern readers with little experience with cursive writing. Many off-line handwritten document image-processing algorithms have recently been proposed in the interest of word recognition and indexing [10], [11], mostly without associating with problem with time series. While handwriting is not a time series, there exist several techniques to convert handwriting to time series. Recent work suggests that this representation may still allow the high precision in indexing historical archives while simplifying the problem from 2-dimentional to 1-dimentional domain [12]

1.2 Existing work on time series retrieval

The explosion of interest in time series indexing in the last decade has been extraordinary, but the majority of the work has focused on the Euclidean distance, which assumes linear mappings between the query and the candidate time series. However, recent work has demonstrated that this similarity model generally does not work well for many real-world problems, where variability in the time axis is always present. This problem of distortion in the time axis can be addressed by Dynamic Time Warping (DTW), a distance measure that has long been known to the speech processing community [13], [14]. This method allows non-linear alignments between the two time series to accommodate sequences that are similar but out of phase. The superiority of DTW over Euclidean distance for classification and indexing has been demonstrated by several authors on a diverse set of domains, including bioinformatics, music retrieval, handwritten document archives, biometrics, chemical engineering, industry, and robotics [3], [4]. Our approach takes this recent work on DTW as its starting point, then fine-tune the algorithm, for a particular domain, and even a particular query, by selectively limiting the amount of warping we allow along various parts of the query. As we will demonstrate, by selectively limiting the amount of warping allowed, we can actually improve the accuracy of DTW, as well as the indexing performance. Due to space limitations, please refer to the DTW background and related work in [15].

2 The Ratanamahatana-Keogh Band (*R-K Band*)

The 'global constraint' of DTW has been almost universally applied to DTW, primarily to prevent unreasonable warping and to speed up its computation. However, surprisingly little research has looked at discovering the best shape and size of the window. Most practitioners simply use one of the well-known bands, e.g., Sakoe-Chiba Band [14] or Itakura Parallelogram [16], proposed in the context of speech recognition several decades ago. In addition, there are several widespread beliefs about DTW, which have been proven to be false by our extensive experiments on wide variety of datasets [17]. The motivation for our work has sparked from this finding that the effect of the window shape/size on accuracy is very substantial, and is strongly domain dependent. Our ideal solution is then to find an optimal band for a given problem that will potentially improve the accuracy. We recently proposed a new representation, the *Ratanamahatana-Keogh Band (R-K Band)* [15], which allows arbitrary shaped constraints.

2.1 A general model of global constraints

To represent a warping window, we can define it in terms of a vector R, $R_i = d$, where $0 \leq d \leq m$, $1 \leq i \leq m$ and R_i is the height above the diagonal in the y direction, as well as the width to the right of the diagonal in the x direction. The classic Euclidean distance can be defined in terms of $R_i = 0; 1 \leq i \leq m;$ only the diagonal path is allowed. More generally, we can define any arbitrary constraint with a vector *R*. Figure 4 illustrates some examples of our *R-K Bands*. An interesting and useful property of our representation is that it also includes the Euclidean distance and classic DTW as special cases

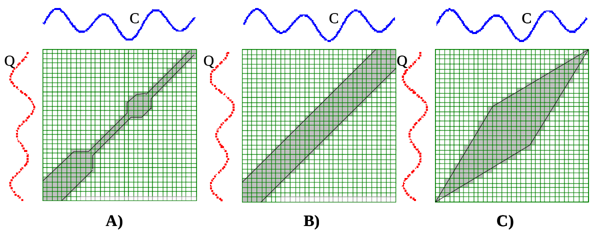

Figure 4. We can use *R* to create arbitrary global constraints. A) We can use *R* to specify all existing global constraints, including the Sakoe-Chiba Band B) and the Itakura Parallelogram C)

. We can exploit the *R-K Bands* for both classification and indexing (query retrieval) problems, depending on the task at hand. In particular,

- for classification, we can use a different *R-K Band* for each class; we denote the band learned for the c^{th} class, as the *R-*K_c *Band*.
- for indexing, we can use *one R-K Band* that maximizes the trade off between efficiency and precision/recall.

Having introduced an *R-K Band*, we can represent any arbitrary warping windows. However, *discovering* the optimal *R-K Band* still remains a question. In some cases, it maybe is possible to manually construct the bands, based on domain knowledge. Unfortunately, our preliminary attempts to manually construct *R-K Bands* met with limited success, even for simple toy problems. Furthermore, since the number of possible *R-K Bands* is exponential, exhaustive search over all possibilities is clearly not an option. We will show how we can *learn* the high-quality bands automatically from data.

2.2 Learning multiple *R-*K_c *Bands* for classification

While it is generally not possible to handcraft accurate *R-K Bands*, it *is* possible to pose the problem as a search problem. Using generic heuristic hill-climbing search techniques, we can perform both forward and backward searches. The forward search starts with the initial Sakoe-Chiba band (uniform) of width 0 (Euclidean), and the backward search starts from the uniform band of the maximum width *m*, above and to the right of the diagonal. The searches are complete when one of the following is true:-No improvement in accuracy can be made; the width of the band reaches *m* for the forward search and 0 (Euclidean) for the backward search or; each section of the band (after recursively cut the portion in half) reaches some threshold. The utility of our *R-*K_c *Bands* for classification has been extensively shown in [18].

2.3 Learning one *R-K Band* for indexing

In addition to creating *R-*K_c *Bands* for classification, we can learn one single *R-K Band* for indexing or query retrieval. The one-band learning algorithm is very similar to the multiple-band learning in the previous section, except that we only maintain one single band representing the whole problem and that we measure the precision/recall instead of the accuracy. Utilizing an *R-K Band* for an indexing problem has been shown in [18] to improve both precision and recall by a wide margin. However, an *R-K Band* needs to be learned from a training data, which may not be practical or available in many circumstances. To resolve this problem, we can build a training data through relevance feedback system, with some help from a user in identifying the positive and negative examples to the system, which is demonstrated in the next section.

3 Relevance Feedback

In text-mining community, relevance feedback is well known to be effective method to improve the query performance [4], [19], [20]. However, there has been much less research in non-text domains. In section 1.1, we have introduced time series as an alternative representation of some multimedia data. This section shows how to incorporate that technique into relevance feedback system using our proposed *R-K Band*.

3.1 Query refinement

Relevance feedback methods attempt to improve performance for a particular informational need by refining the query, based on the user's reaction to the initial retrieved documents/objects. Working with time series retrieval is rather similar to the text retrieval; a user can draw or provide an example of a query and retrieve the set of best matches' retrieval of data. Once the user ranks each of the results, a query refinement is performed such that a better-quality query is produced for the next retrieval round. In our system, the user is asked to rank each result in a 4-point scale, based on relevance to their informational needs, which are converted into appropriate weights being used in query refinement processes (averaging the positive results with current query). However, averaging a collection of time series that are not perfectly time-aligned is non-trivial and DTW is needed [21]; Hierarchically, each pair of time series are averaged based on weights and warped alignments.

3.2 *R-K Band* learning in relevance feedback

We will empirically demonstrate that our proposed *R-K* Band combined with the query refinement can improve precision and recall of retrieval. Table 1 shows our algorithm.

Table 1. *R-K Band* learning with Relevance Feedback algorithm

```
Algorithm RelFeedback(initial_query)

1.  Repeat until all rankings are positive.
2.  Show the 10 best matches to the current query to the user.
3.  Let the user rank how relevant each result is.
4.  According to the ranking, accumulatively build the training set;
    positive result → class 1, negative result → class 2.
5.  Learn a single envelope that represents the given training data.
6.  Generate a new query, by averaging (with DTW) the positive results
    with the current query according to their weights.
7.  end;
```

In the first iteration, given a query, the system uses the initial *R-K Band* (size 0), retrieves the 10 nearest neighbors, and then shows them to the user (line 1). After ranking, the positive and negative responses are noted and collected as a training data (lines 3-4). The algorithm uses this training data to learn an *R-K Band* that best represents the positive objects in the training set while being able to correctly differentiate the positive from the negative instances (line 5). The training data will be accumulated during each round, thus producing progressively finer results. The process is complete when only positive feedbacks are returned or the user abandons the task. In our experiments, we consider 3 multimedia datasets, shown in Table 2, in our relevance feedback technique (complete datasets details are available in [18]). To evaluate our framework, we measure the precision and recall for each round of the relevance feedback retrieval. Since we only return the 10 best matches to the user and we would like to measure the precision at all recall levels, we purposely leave only 10 relevance objects of interest in all the databases.

Table 2. Datasets used in the relevant feedback experiment

Dataset	Number of Classes	Time Series Length	Number of items
Gun	2	150	200
Leaf	6	350	442
Wordspotting	4	100	272

We then measure the performance of our relevance feedback system with the precision-recall plot from each round of iteration. Figure 5 shows the precision-recall curves of the three datasets for the first five iterations of relevance feedback. Our experiments illustrates that each iteration results in significant improvement in both precision and recall.

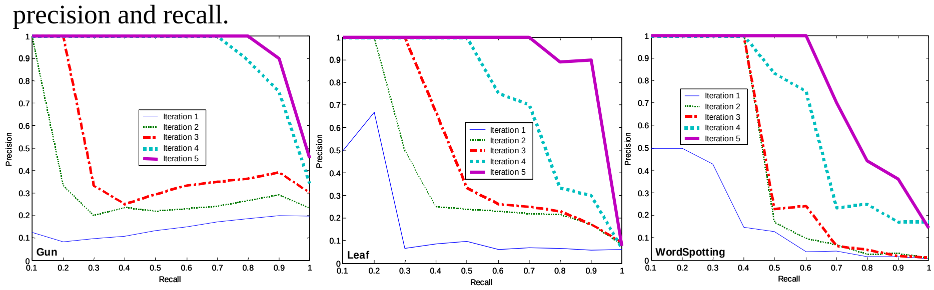

Figure 5. The precision-recall plot for the 3 datasets with 5 iterations of relevance feedback

4 Conclusion

We have introduced a framework for both classification and time series retrieval. The Ratanamahatana-Keogh Band (R-K Band) allows for any arbitrary shape of the warping window in DTW calculation. With our extensive evaluation, we have shown that our framework incorporated into relevance feedback can reduce the error rate in classification, and improve the precision at all recall levels in video and image retrieval.

Acknowledgments: This research was partly funded by the NSF under grant IIS-0237918.

References

1. Agrawal, R., Lin, K. I., Sawhney, H. S., & Shim, K. (1995). Fast similarity search in the presence of noise, scaling, and translation in times-series databases. VLDB, pp. 490-501.
2. Faloutsos, C., Ranganathan, M., & Manolopoulos, Y. (1994). Fast subsequence matching in time-series databases. In Proc. ACM SIGMOD Conf., Minneapolis. pp. 419-429.
3. Aach, J. & Church, G. (2001). Aligning gene expression time series with time warping algs. Bioinformatics. 17: 495-508.
4. Bar-Joseph. Z., Gerber, G., Gifford, D., Jaakkola T & Simon. I. (2002). A new approach to analyzing gene expression time series data. RECOMB, pp. 39-48.
5. Kadous, M.W. (1999). Learning comprehensible descriptions of multivariate time series. ICML: 454-463.
6. Schmill, M., Oates, T., & Cohen, P. (1999). Learned models for continuous planning. In 7th International Workshop on Artificial Intelligence and Statistics.
7. Keogh, E. and Kasetty, S. (2002). On the Need for Time Series Data Mining Benchmarks: A Survey and Empirical Demonstration. In the 8th ACM SIGKDD. pp. 102-111.
8. Gandhi, A. (2002). Content-based Image Retrieval: Plant species Identification. MS Thesis, Oregon State U.
9. Jain, A.K. & Namboodiri, A.M. (2003). Indexing and Retrie-val of On-line Handwritten Documents. ICDAR: 655-659.
10. Kavallieratou, E., Dromazou, N., Fakotakis, N., & Kokkinakis, G. (2003). An Integrated System for Handwritten Document Image Processing. IJPRAI, (17)4, pp. 617-36.
11. Tomai, C.I., Zhang, B., & Govindaraju, V. (2002). Transcript Mapping for Historic Handwritten Document Images. IWFHR'02.
12. Rath, T. & Manmatha, R. (2003). Word image matching using dynamic time warping. CVPR: II, 521-27.
13. Kruskall, J. B. & Liberman, M. (1983). The symmetric time warping algorithm: From continuous to discrete. In Time Warps, String Edits and Macromolecules.
14. Sakoe, H. & Chiba, S. (1978). Dynamic programming algorithm optimization for spoken word recognition. IEEE Trans. Acous., Speech, and Signal Proc., ASSP-26: 43-49.
15. Ratanamahatana, C.A. & Keogh, E. (2004). Making time-series Classification More Accurate Using Learned Constraints. SDM International Conference, pp. 11-22.
16. Itakura, F. (1975). Minimum prediction residual principle applied to speech recognition. IEEE Trans. Acoustics, Speech, and Signal Proc., Vol. ASSP-23, pp. 52-72
17. Ratanamahatana, C.A. & Keogh, E. Three Myths about Dynamic Time Warping Data Mining. SIAM Int'l Conference on Data Mining (SDM), Newport Beach, CA, (2005).
18. Ratanamahatana, C.A. (2005). Improving Efficiency and Effectiveness of Dynamic Time Warping in Large Time Series Databases. Ph.D. Dissertation, Univ. of Calif., Riverside.
19. Rocchio, J.J. (1971). Relevance Feedback in Information Retrieval. The SMART Retrieval System: Experiments in Automatic Document Processing. pp. 313-323.
20. Salton, G., & Buckley, C. (1990). Improving retrieval performance by relevance feedback. JASIS41(4),.288-297.
21. Gupta, L., Molfese, D., Tammana, R., and Simos, P. (1996). Nonlinear Alignment and Averaging for Estimating the Evoked Potential. IEEE Trans. on Biomed Eng(43)4.