PERCEPTUAL INDEXING OF MULTIVARIATE TIME SERIES

Aleksandra Mojsilović
IBM T. J. Watson Research Center, Yorktown Heights, NY, email: alekand@us.ibm.com

ABSTRACT

We consider the problem of deriving compressed perceptual representation of multivariate time series and using it for efficient indexing and similarity search. Our algorithm is based on the identification of perceptual skeletons in multidimensional space and the use of these "simplifications" in similarity measurements. We illustrate the performance of the algorithm in a financial modeling application. Our results indicate that the skeleton representation outperforms the traditional approaches and is robust enough to be used even with the simplest distance metrics.

Index Terms— indexing, time-series, matching

1. INTRODUCTION

A large portion of the digitally stored data are in form of a time series, and the ability to efficiently search and organize such data is of growing importance in many applications. As a result, a significant effort has been directed towards developing methods that will enable computers to assist users in performing tasks such as: "find companies with similar stock prices", "find portfolios that behave similarly", "find products with similar sell cycles", "cluster users with similar credit card utilization", "search for music", etc. Prior work includes the application of the Discrete Fourier Transform, Discrete Wavelet Transform, Principal Component Analysis or Linear Predictive Coding cepstrum representation to reduce sequences into points in low dimensional space and the use of the Euclidean distance between two sequences as a measure of similarity [1]-[5]. However, there are many similarity queries where Euclidean distances fail to capture the notion of similarity. A concept that two series should be considered similar if they have enough non-overlapping time-ordered pairs of similar subsequences has been explored in [8]. Rafiei et al used a set of linear transformations on the Fourier series representation of a sequence as a basis for similarity measurement [6], while Yi et al used the time warping distance [7]. A special class of problems is the analysis and classification of multivariate time series. Examples include electroencephalograms (where measurements are recorded up to dozens of channels), weather data (with daily measurements of temperature, humidity, atmospheric pressure and wind), and stock market portfolios (with multiple stocks tracked over a period of time). In [13] Taniguchi showed that similarities and differences between multivariate stationary time series can be characterized in terms of the structure of the covariance or spectral matrices. Huan et al proposed using the library of smooth localized complex exponentials to extract computationally efficient features of nonstationary time series [14]. A separate area of research focused on the design of feature sets that will allow for more effective and "perceptually tuned" representation of time series based on the extraction of key features, event detection, and extraction of important points [9]-[12]. These techniques are especially interesting as they attempt to capture the notion of similarity from the perspective of human observer. However, most of these perceptual techniques have difficulties handling multivariate data. This paper considers a problem of deriving a simple, compressed perceptual representation of multivariate time series, and using it as a basis for efficient indexing and search.

2. OVERVIEW OF THE WORK

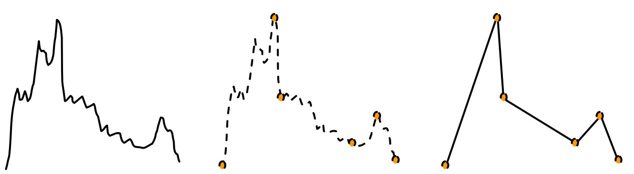
Algorithms that attempt to capture some elements of human perception have often showed excellent results in many applications. When assessing similarity, humans mine visual data extensively to construct a representation that captures the most important aspects of a signal, the nature of the application and the task that needs to be achieved. For example, humans are very good at constructing different representations of an object, simplifying them by "picking" the most important characteristics, and using these "simplifications" to derive similarity judgments. Therefore, at the core of any similarity task is the computation of a *perceptual skeleton*, a set of points we "care about", and using them in the matching task. Although such a process is difficult to generalize, by including its key steps into a matching algorithm one can greatly improve the accuracy or relevance of retrieved results. Our methodology attempts to follow this process. We propose a framework for similarity measurement of time domain signals with multiple attributes. The first step in our methodology involves transforming signals into a space with a metric (*constructing the representation*), where we can perform operations such as measuring distances between different points of the signal or identifying local maxima. We then compute the skeleton of a signal as a set of perceptually important points in that space (*constructing the skeleton*), and use a distance between the two skeletons as a similarity measure (*similarity* *measurement*).

3. PRELIMINARIES

Let us consider two discrete time domain signals, $\mathbf{X}=[\mathbf{x}(t_1),...,\mathbf{x}(t_{N_x})]$, and $\mathbf{Y}=[\mathbf{y}(t_1),...,\mathbf{y}(t_{N_y})]$, of length N_x, and N_y, respectively. Each time instance is described with M attributes, $\mathbf{x}(t)=[x_1(t),\ldots,x_M(t)]$, and $\mathbf{y}(t)=[y_1(t),\ldots,y_M(t)]$. Usually the attribute vectors represent different measurements, which are often either strongly correlated, or include features that are distinctly different in nature so that a distance metric between two attribute vectors cannot be defined naturally. Therefore, as a first step we apply a de-correlating transform $F(\cdot)$ and project $\mathbf{X}$ and $\mathbf{Y}$ onto a *K*-dimensional metric space, S

$$\mathbf{F_X}=F(\mathbf{X})=[\mathbf{f_X}(t_1),\ldots,\ \mathbf{f_X}(t_{N_x})],$$
$$\mathbf{F_Y}=F(\mathbf{Y})=[\mathbf{f_Y}(t_1),\ldots,\mathbf{f_Y}(t_{N_y})] \qquad (1)$$

where, $\mathbf{f}(t)=[f_1(t),\ldots,\ f_K(t)]$ and $K\le M$. We will also assume that S is a normed linear space with a norm, $\|\cdot\|$, and metric $d(\mathbf{f_X},\mathbf{f_Y})=\|\mathbf{f_X}-\mathbf{f_Y}\|$ defined by the norm. Note that the goal of the mapping is not dimensionality reduction (although this is a useful step when dealing with highly correlated variables), but the projection of a signal into a space where a metric can be defined more naturally. This metric will then constitute a *local similarity metric*, used to identify perceptually skeletons, compute the compression rate and construct a *global similarity metric* (i.e. a true similarity distance between the two signals).



a) b) c)

Figure 1: An illustration of perceptually important points and perceptual skeletons for a one-dimensional signal. a) Original signal, b) perceptually important points, and c) perceptual skeleton obtained by connecting the PIPs.

4. COMPUTING PERCEPTUAL SKELETONS

A body of research in cognitive psychology indicates that humans and animals depend on "landmarks" and "simplifications" in organizing their spatial memory. A subject asked to look at Fig. 1a and duplicate the picture, will typically memorize only the key turning points, as in Fig. 1b, and then recreate the picture by connecting these few points, Fig. 1c. This idea of perceptually important features has been explored in a variety of applications. One of the first uses of this concept was in reducing a number of points required to represent a line in cartoon making [15]. Similar ideas have also been explored independently in [11], [12], [16]. Here, we define a *perceptually important point* (PIP) as a local maximum of the transformed signal, $\mathbf{F}$ (depending on the nature of the problem, one can use maxima of different orders). At the coarsest level, each point in $\mathbf{F}$ represents a PIP. The idea behind the perceptual skeletons is to discard minor fluctuations and keep only major maxima. One possible PIP identification procedure for one-dimensional signals is described in [16]. Here we refine the procedure and extend it to handle multi-dimensional feature representations. We start with the transformed signal $\mathbf{F}=[\mathbf{f}(t_1),\ldots,\ \mathbf{f}(t_N)]$, and select the first and the last point as the first two PIPs. Every next PIP is then identified as a point with the maximum distance to its two adjacent PIPs. Fig. 2a illustrates this procedure in one dimensional case, while Fig. 2b represents a generalization to multiple dimensions. The PIP identification procedure can be then described as follows:

$$\begin{aligned}\mathrm{PIP}_1&=[1,\mathbf{f}(t_1)]=[z_1(1),\ z_2(1),\ldots,z_{K+1}(1)],\\ \mathrm{PIP}_2&=[N,\mathbf{f}(t_N)]=[z_1(N),\ z_2(N),\ldots,z_{K+1}(N)],\\ \mathrm{PIP}_3&=[i,\mathbf{f}(t_i)]=[z_1(i),\ z_2(i),\ldots,z_{K+1}(i)],\end{aligned} \qquad (2)$$

where $i=\arg\max_i d(\mathbf{f}(t_i),\mathbf{fn}(t_i))$, and

$$\mathbf{fn}(t_i)=[tn(i),\ fn_1(t_i),\ fn_2(t_i),\ldots,\ fn_K(t_i)]=$$
$$=[zn_1(i),\ zn_2(i),\ldots,zn_{K+1}(i)]$$

is obtained by projecting the point $\mathbf{f}(t_i)$ onto a line, $\ell_{\mathrm{PIP_1,PIP_2}}$, which connects the two neighboring PIPs (as illustrated in Fig. 2). A line in *K+1*-dimensional space can be represented as

$$z_i=m_{i-1}z_{i-1}+n_{i-1},\ i=2,\ldots,K+1,$$

hence, the line connecting PIP_1 and PIP_2 is defined by:

$$m_{i-1}=\frac{z_i(N)-z_i(1)}{z_{i-1}(N)-z_{i-1}(1)},$$

$$n_{i-1}=\frac{z_i(1)z_{i-1}(N)-z_i(N)z_{i-1}(1)}{z_{i-1}(N)-z_{i-1}(1)},\ i=2,\ldots,K+1$$

From now on we will assume L^2 norm to be the local similarity metric in the space – in that case, for every point $\mathbf{f}(t_i)$, $\mathbf{fn}(t_i)$ can be found by maximizing:

$$D=d(\mathbf{f}(t_i),\mathbf{fn}(t_i))=\sum_{j=1}^{K+1}(z_j(i)-zn_j(i))^2,$$

subject to: $zn_j(i)=m_{i-1}zn_{j-1}(i)+n_{i-1},\ i=2,\ldots,K+1$

(i.e. subject to: $\mathbf{fn}(t_i)\in\ell_{\mathrm{PIP_1,PIP_2}}$). Using Lagrange multipliers ($\lambda_1,\ldots,\lambda_K$) to solve this problem, we obtain $\mathbf{fn}(t_i)$ as a solution to the following system of equations

$$zn_1(i)+\frac{1}{2}\lambda_1 m_1=z_1(i),$$
$$zn_j(i)-\frac{1}{2}\lambda_{j-1}+\frac{1}{2}\lambda_j m_j=z_j(i),\ j=2,\ldots,K$$
$$zn_{K+1}(i)-\frac{1}{2}\lambda_K=z_{K+1}(i),$$

$$zn_{j+1}(i) - m_j zn_j(i) = n_j,\ \ j = 1,\ldots,K$$

The identification process continues until a distortion measure (as defined below) is satisfied, or the number of PIPs is equal to the length of the sequence. The local similarity measure d can be also used as a distortion measure. Given the transformed sequence $\mathbf{F}$ and the perceptual skeleton, $\mathbf{F}_s = [\mathbf{f}_s(t_1),\ldots,\mathbf{f}_s(t_N)]$, obtained by connecting the PIPs along the original time axis, the distortion rate, dr, can be computed as:

$$dr = \frac{1}{N}\sum_{i=1}^{N} d(\mathbf{f}(t_i), \mathbf{f}_s(t_i))$$

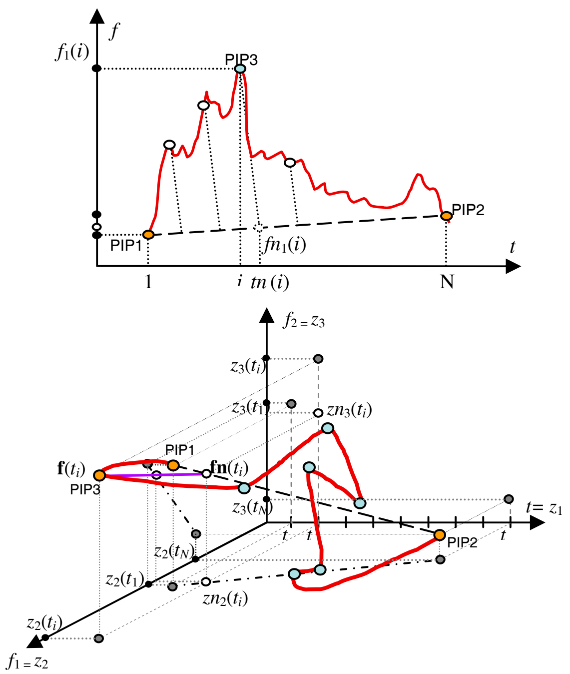


Figure 2: *Identification of perceptually important points for: one dimensional case (top), and multivariate time series (bottom).*

5. SIMILARITY MEASUREMENTS

Once the signals $\mathbf{X}$ and $\mathbf{Y}$ are reduced to their perceptual skeletons $\mathbf{F}_s^{\mathbf{X}}$ and $\mathbf{F}_s^{\mathbf{Y}}$ the final step is to compute the similarity between the simplified representations. We will first consider the local similarity metric, d, as a global distance measure. However, as it is often reported that Minkowski-based metrics have drawbacks in comparing time series. Therefore, we will also consider multivariate dynamic time warping (DTW) as an alternative measure [7]. We start with the perceptual skeletons $[\mathbf{f}_s^{\mathbf{X}}(t_1),\ldots,\ \mathbf{f}_s^{\mathbf{X}}(t_{N_x})]$ and $[\mathbf{f}_s^{\mathbf{Y}}(t_1),\ldots,\ \mathbf{f}_s^{\mathbf{Y}}(t_{N_y})]$, where N_x and N_y are the number of points in each skeleton, respectively. To compute the similarity measure between the skeletons, we first construct an $N_x \times N_y$ matrix M, where $M(i,j) = d(\mathbf{f}_s^{\mathbf{X}}(t_i), \mathbf{f}_s^{\mathbf{Y}}(t_j))$, and d is the local similarity metric (as defined in Section 3). The *warping path*, $W = w_1, w_2,\ldots, w_L$, where $w_l = (i,j)_l$ is a contiguous set of matrix elements that defines a mapping between $\mathbf{F}_s^{\mathbf{X}}$ and $\mathbf{F}_s^{\mathbf{Y}}$, subject to: *boundary conditions* $w_1 = (1,1)$ and $w_L = (n_x, n_y)$, *continuity constraint* $w_k = (a,b) \Rightarrow w_{k-1} = (a',b')$, where $a - a' \le 1$ and $b - b' \le 1$, and *monotonicity constraint* $a - a' \ge 0$ and $b - b' \ge 0$. As there are many warping paths that satisfy these conditions, we are interested in finding the path that minimizes the warping cost

$$DTW(\mathbf{F}_s^{\mathbf{X}}, \mathbf{F}_s^{\mathbf{Y}}) = \min_W \sqrt{\sum_{l=1}^{L} M(w_l)}$$

6. EXPERIMENTS AND RESULTS

We will illustrate the application and the performance of the algorithm in a financial modeling application, and use the dataset consisting of 1996-2006 daily stock prices for the DOW Jones Industrial (DJI) index. The index includes 32 stocks. We first illustrate the application of the algorithm using the simple one dimensional case. We select a stock and time interval of interest and search for the stocks that performed similarly in the same time period. Fig. 3 shows a retrieval example using the proposed method and the Euclidean distance between the original signals.

Figure 3: *A retrieval example for one dimensional case. The query is the stock price series for American Express, in a 3 month period starting on 11/14/2005. Using the skeleton representation we obtain JP Morgan as the closest match (both using the Euclidean distance and DTW). The closest match using the Euclidean distance on the original time series is Hewlett Packard.*

We will now demonstrate the application of the method in a multi-dimensional setting by considering the following model of the stock market. We will assume a market with Q assets. Market vectors $\mathbf{p}(t) = [p_1(t),\ldots, p_Q(t)]$ and $\mathbf{r}(t) = [r_1(t),\ldots, r_Q(t)]$ are vectors of nonnegative numbers representing asset daily prices and returns (price relatives) for every trading day. Let us assume the following simple sequential "momentum" investment strategy. An investor starts investing at time T_o and rebalances her portfolio every T_r days. The investor can invests all her wealth into only one stock. Let S_o denote investor's initial capital. Then, at the end of the first trading period the investor's wealth, S_1, and return, R_1, become:

$$S_1 = \prod_{t=To+1}^{To+T_r} S_o r_i(t),\quad R_1 = S_1 / S_o = \prod_{t=To+1}^{To+T_r} r_i(t)$$

where i is the index of the asset selected for investment. In order to select the investment for the next trading period, the investor will consider the evolution of the market over T_h

days prior to the decision time, which is represented by a sequence of price vectors $\mathbf{P}(t) = [\mathbf{p}(t - T_h), \ldots, \mathbf{p}(t-1)]$. The investor will analyze the stock market history, find a period when the market behaved similarly to the current one, identify the asset that had the highest return in the given period and select that asset as the new investment. In other words, at the beginning of every trading period, t_i, the investor finds the index of the new investment as

$$ind(i) = \underset{t_j = T_h+1,\ldots,t_i-1}{\arg\min} \; D(\mathbf{P}(t_i), \mathbf{P}(t_j))$$

where $T_h + 1, \ldots, t_i - 1$ is the "market history" prior to t_i. The investor's return after N trading periods then becomes

$$R = S_N / S_o = \prod_{n=1}^{N} R_n = \prod_{n=1}^{N} \prod_{t=T_o+(n-1)Tr+1}^{T_o+nT_r} r_{ind(T_o+(n-1)T_r)}(t)$$

The sequence of price vectors $\mathbf{P}(t)$ is a Q-dimensional time series, where each point represents a market vector at time t. Thus, we will use our algorithm to find the most similar past market conditions, and will evaluate the performance of our method by comparing the achieved total return R, to the returns obtained by using Euclidian distance (ED) and dynamic time warping (DTW) as similarity metrics between the original signals. We will also compare the performance of the perceptual skeletons with DTW as similarity metric (PS+DTW), with the Euclidean distance as similarity metric (PS+ED). Instead of the distortion rate, we control the quality of the representation via the parameter $SL_{\min}$, which defines the minim length of a segment between two PIPs. Results for different choices of $(T_r, T_h, SL_{\min})$ are given in Table 1.

TABLE 1: THE RESULTS OF THE MOMENTUM INVESTMENT STRATEGY FOR DIFFERENT CHOICES OF REBALANCING PERIOD, *Tr*, MARKET HISTORY, *Th*, AND *SL*min. WE USE THE 1996-2006 PRICES FOR DJI INDEX, THUS *Q*=32 IN OUR CASE.

(T_r,T_h,SL_{min})	PS+DWT	PS+ED	DWT	ED
(40,90,10)	2.28	2.28	1.82	2.09
(20,90,10)	2.01	1.92	1.82	1.34
(15,90,10)	3.18	3.18	1.85	3.09
(90,90,5)	2.17	2.17	1.26	2.17
(90,90,15)	2.36	2.36	1.26	2.17
(90,90,20)	1.81	1.81	1.26	2.17
(120,120,3)	1.57	1.57	1.96	1.57
(120,120,5)	2.36	2.36	1.96	1.57
(120,120,15)	2.60	2.60	1.96	1.57
(120,120,20)	2.17	2.17	1.96	1.57

The skeleton based representation clearly outperforms the other methods. Furthermore, when using perceptual skeletons, both DTW and ED generate the same returns indicating that the perceptual representation is robust with respect to the selection of distance measures. We also observe how the performance of the skeleton representations depends on the compression factor and deteriorates as the representation becomes to coarse (large $SL_{\min}$, resulting in large distortion rates), or when the simplification is insufficient (too small $SL_{\min}$, yielding a signal representation that is similar to the original signal), and it is of interest to study these relationships in more detail.

REFERENCES

[1] R. Agrawal, C. Faloutsos, and A. Swami, "Efficient similarity search in sequence databases", *Proc. of the 4th Int. Conf. on Foundations of Data Organization and Algorithms*, pp. 69-84, Chicago, October 1993.

[2] C. Faloutsos, M. Ranganathan, and Y. Manolopoulos, "Fast subsequence matching in time-series databases", *Proc. of the ACM SIGMOD Int. Conf. on Management of Data*, pp. 419-429, Minneapolis, May 1994.

[3] F. Chan, A. Fu, "Efficient Time Series Matching by Wavelets", *15th IEEE International Conference on Data Engineering*, Sydney, Australia, March 23-26, 1999, pp. 126-133.

[4] M. Gavrilov, D. Anguelov, P. Indyk, and R. Motwani. "Mining the stock market: Which measure is best?" *Proc. of the KDD*, pp. 487-496, 2000.

[5] K. Kalpakis, D. Gada, V. Puttagunta, "Distance measures for effective clustering of ARIMA time—series," *Technical Report TR-CS-01-14*, CSEE, UMBC, 2001.

[6] D. Rafiei and A.O. Mendelzon, "Similarity-based queries for time series data," *Proc. ACM SIGMOD*, Tucson, May 1997.

[7] B. Yi, K. Jagadish, and C. Faloutsos, "Efficient retrieval of similar time sequences under time warping," *Proc. Int. Conf. on Data Engineering*, pp 23-27, 1998.

[8] R. Agrawal, K.-I. Lin, H.S. Sawhney, and K. Shim, "Fast similarity search in the presence of noise, scaling, and translation in time-series databases," Proc. of the 21st VLDB conference, Zurich, Switzerland, 1995.

[9] V. Guralnik, and J. Srivastava, "Event detection from time series data," *Proceedings of the fifth ACM SIGKDD Int. Conf. on Knowledge discovery and data mining*, pp. 33-42, 1999.

[10] M. Last, Y. Klein, A. Kandel, "Knowledge discovery in time Series Databases," *IEEE Transactions on Systems, Man, and Cybernetics*, Part B, vol. 31, no. 1, pp. 160-169, 2001.

[11] K. B. Pratt, and E. Fink, "Search for patterns in compressed time series," *Int. Journal of Image and Graphics*, vol. 2, no. 1, pp. 89-106, 2002.

[12] C-S. Perng, H. Wang, S. Zhang, D. S. Parker, "Landmarks: A new model for similarity-based pattern querying in time series databases," *Proc. of the 16 Int. Conf. on Data Engineering* (ICDE), San Diego, CA, February 2000.

[13] M. Taniguchi, "Discrimination and clustering for multivariate time series," *J. of the American Stat. Association*, March, 1998.

[14] H-Y. Huang, H. Ombao, and D. S. Stoffer, "Discrimination and classification of nonstationary time series esing the SLEX Model," *Journal of the American Statistical Association*, vol. 99, no. 467, pp. 763-774, September 2004.

[15] D. Douglas, and T. Peucker, "Algorithms for the reduction of the number of points required to represent a digitized line or its caricature", *The Canadian Cartographer*, vol. 10, no, 2, pp. 112-122, 1973.

[16] T. C. Fu, F. L. Chung, R. Luk, and C. M. Ng, "Financial Time Series Indexing Based on Low Resolution Clustering," *4th IEEE Int. Conf. on Data Mining* , Brighton, UK. Nov 2004.