Application of time series techniques to data mining and analysis of spatial patterns in 3D images

Qiang Wang Despina Kontos Guo Li Vasileios Megalooikonomou

Department of Computer and Information Sciences, Temple University, Philadelphia, PA

ABSTRACT

Analysis of spatial patterns in images can provide valuable information in many application domains, such as in geography, meteorology and medicine. We propose to apply techniques from the time series domain to analyze the spatial patterns extracted from 3D images. After traversing an image using a space-filling curve we discover discriminative patterns by analyzing the spatial sequence in the transformed domain. Because of the similarity of the sequences with time series we propose the use of existing time series similarity analysis techniques, including Euclidean distance and dimensionality reduction techniques such as Singular Value Decomposition and Piecewise Aggregate Approximation, for further analysis of the spatial patterns. As a case study, we analyze an fMRI dataset. Experimental results verify that the discovered spatial patterns have strong discriminative power among different classes and the overall accuracy for clustering and similarity retrieval is above 90% and as high as 100% for certain experimental settings.

1. INTRODUCTION

Spatial patterns play a crucial role in a wide range of scientific and engineering problems where data often comes as images. For human brain images, several approaches have been used to detect discriminative patterns [1], but there is usually lack of methods to automatically classify such patterns and quantitatively measure levels of their similarity.

Kontos et al [2] proposed a new technique to detect and analyze discriminative patterns in 3D images. This technique is based on an initial linear mapping of the multidimensional space using the Hilbert space-filling curve [3]. This space-filling curve has been proved optimal in preserving the locality and clustering properties of data after the domain transformation [4]. Statistical tests of significance are then applied to groups of points (that correspond to voxels) in the transformed domain to detect discriminative patterns. Grouping the points as part of a binning process reduces the multiple comparison problem observed in voxel-based analysis techniques [5]. The significance of the discriminative patterns extracted by this new technique has been shown through classification experiments using neural networks.

However, considering the high dimensionality of the discovered patterns, it is still difficult to do further analysis such as clustering, classification and similarity searches. Noticing that the voxel values of discriminative patterns in 1D are arranged in a certain spatial order and can actually be viewed as a time series (or sequence), we propose to apply existing analysis techniques already available in the time series domain. This idea is also motivated by the fact that there is an enormous wealth of time series analysis methods and associated dimensionality reduction techniques that can help overcome the difficulties mentioned above.

In the time series domain, the dimensionality reduction has been considered as one of the most important issues since spatial access methods perform well only when the number of dimensions is low (typically below 15). The solution to extract a signature from each sequence and to index the signature space was originally proposed by Faloutsos et al [6]. To guarantee completeness (i.e., no false dismissals) the admissibility criterion that the distance function used in the signature space must underestimate the true distance measure (bounding lemma) was also proposed [6]. Generally speaking, there are two groups of techniques to reduce the dimensionality of time series: those based on global information, such as Singular Value Decomposition (SVD) [7]; and those based on local information, such as Discrete Fourier Transform (DFT) [8] and Piecewise Aggregate Approximation (PAA) [7]. A survey of recent methods can be found in [9].

Here as a case study, we analyze a functional MRI¹ dataset obtained from a study of Alzheimer's disease. We apply several time series analysis techniques to the discriminative fMRI activation patterns detected using space-filling curves and statistical tests.

2. Methodology

¹ Functional-Magnetic Resonance Imaging: shows physiological activity in brain

As proposed in [2], we extract discriminative fMRI activation patterns by traversing the 3D space of fMRI images using the Hilbert space-filling curve and applying a statistical test of significance on groups of voxels (points) in the transformed linear domain. Since Hilbert space-filling curve preserves the locality in the original space, the neighboring voxels in the 3D space are also close to each other in the linear domain. Figure 1 shows an example of the mapping. After the mapping, we apply



Figure 1. (a) A 2D slice of a 3D fMRI map, and (b) its traversal using the Hilbert space-filling curve.

statistical analysis (t-test) in the linear domain. In order to reduce the computational complexity and the effect of the multiple comparison problem, we employ an intermediate step of binning: K neighboring voxels are grouped together and their mean value is taken as the representative attribute. The voxel activation of a specific bin is considered to differ class-wise substantially when the statistical significance of the divergence is above a certain predefined threshold θ .

Here, we propose to concatenate the values of voxels in all the significant bins in the traversal order of the filling curve and obtain a spatial sequence which contains discriminative information about the original 3D volume. Except for the spatial characteristic, this sequence is very similar to a time series. A time series is a sequence (ordered collection) of real values, $X = x_1, x_2,..., x_n$, where *n* can be very large.

After mapping the discriminative patterns of a 3D volume to a time series, we can further analyze the patterns using time series analysis techniques. Since similarity calculation is the foundation of any further investigation in time series domain, we need to define a distance function.

The most popular distance metric is the Euclidean distance: for a target series X and a query series Q, both of length n, the Euclidean distance is defined as:

$$D(X,Q) = \sqrt{\sum_{i=1}^{n} (q_i - x_i)^2}$$

Obviously, the simplest way of calculating the similarity (or distance) between time series is the plain Euclidean method, which computes the Euclidean distance directly, i.e., on the original series. For a small dataset this can be very useful, though, for large data sets the efficiency may become a problem. Since the time complexity is $O(N^*n)$, where *n* is the number of features that need to be

represented for each time series and N is the number of time series in the dataset.

In order to compute the similarity and index efficiently while keeping the accuracy not significantly affected, many techniques of dimensionality reduction have been suggested, such as DFT [8], SVD, PAA [7], etc. The main idea of most of these techniques is to extract only a few representative coefficients for representing the original time sequences through a certain transformation; in most cases, the coefficients that preserve most of the energy of the original sequences are used. In the time series domain, an ideal transformation should guarantee no false dismissals i.e., the distance between a pair of time series should not be enlarged after the transformation. This requirement can be written as:

$D(F(X), F(Q)) \leq D(X, Q)$

where F(X), F(Q) are the representations of time series X and Q after the transformation.

Discrete Fourier Transform (DFT) and Discrete Wavelet Transform (DWT) have been shown very useful in digital image processing, and they also have been utilized in time series analysis [10]. These two transforms are very similar and the difference between them is that DFT maps a one dimensional time domain discrete function into a representation in frequency domain while the wavelet transform maps it into a representation that allows localization in both time and frequency domains.

Two other popular techniques for the dimensionality reduction are Piecewise Aggregate Approximation and Singular Vector Decomposition. The basic idea of Approximation Piecewise Aggregate (PAA) is straightforward: Given a time series $X = x_1, x_2, ..., x_n$, we segment it into k (k << n) equi-length parts. By recording the mean value of each part, we get a new sequence Y = $\{y_1, y_2, \dots, y_k\}$, whose dimensionality (k) is much lower than that of the original time series. There are several variations of this method, such as the Adaptive Piecewise Constant Approximation (APCA) [11], which can treat parts with different lengths.

While PAA is a technique based on local information (single time series), Singular Vector Decomposition (SVD) is based on the global information of the whole dataset: The entire dataset is examined and rotated in such a way that the first axes has the maximum possible variance, the second axes has the second maximum possible variance orthogonal to the first, the third axes has the third maximum possible variance orthogonal to the first two, etc. Since most variance information is recorded in the first several (k) dimensions after the transformation, it is possible to keep only the first k values of a transformed time series without losing much information.

In the following section, we present the experimental results of a case study using both the plain Euclidean method and dimensionality reduction techniques.

3. RESULTS

3.1. Data preparation

The original dataset consists of 3D activation maps of 9 controls and 9 Alzheimer's disease patients on a categoryexemplar word pair. After certain spatial normalization, we map each 3D map into 1D using Hilbert space-filling curve. The t-test is then employed for detecting discriminative voxels. To test the effect of the threshold, we take two different values for θ : 0.05 and 0.01. With θ =0.05, we get a time series of length 10,000 for each subject and with θ =0.01 we get a time series of length 1,500. Finally, we create two datasets: TS01 (θ =0.01) with 18 time series each of length 1,500 and TS05 (θ =0.05) with 18 time series each of length 10,000.

To avoid the potential effects of scaling and shifting in analysis, before we actually perform any experiment, the datasets are preprocessed using Z normalization. That is, each time series X is normalized as: $X = (X - \overline{X}) / \sigma(X)$ where \overline{X} is the mean value of X and $\sigma(X)$ is its standard deviation. Figure 2 shows an example from each dataset.



Figure 2. Examples of time series: (a) from TS01, (b) from TS05.

3.2. Time Series Analysis

In order to evaluate the discriminative power of the extracted patterns, we first apply the plain Euclidean method which uses the Euclidean distance on the original (full-length) time series. Due to the high dimensionality (1,500 and 10,000), the plain Euclidean method is computationally expensive. In order to improve the computational efficiency, we utilize the following dimensionality reduction techniques introduced in the previous section: DFT, DWT, SVD and PAA. Based on the distance (dissimilarity) calculated with these different methods, we perform experiments on clustering and nearest-neighbor retrieval. To make the results more robust, for dimensionality reduction; 5 and 10. (The dimensionality is doubled for DWT).

3.2.1 Clustering experiments

For time series clustering, the PAM (partitioning around medoids) algorithm was used on both datasets, with time series distance calculated based on different methods.

At the same time, in order to evaluate the accuracy and quality of clustering, the following cluster similarity metric was used: Given two clusterings $G=G_1,G_2,...,G_K$ (the groundtruth), and $A = A_1,A_2,...A_k$ (clustering result by a certain clustering method), the clustering accuracy is evaluated with the cluster similarity defined as:

$$Sim(G, A) = \left(\sum_{i} \max_{j} Sim(G_i, A_j)\right) / k$$

where Sim(G_i, A_j) = $(2 | G_i \cap A_j |) / (|G_i| + |A_j|)$.

Obviously, the value ranges between 0 and 1, with 1 standing for a perfect clustering result. More details about this metric can be found in [12].

Table 1 displays the actual experimental results. Due to the stochastic nature of PAM algorithm, each value in the table is actually the mean of 10 experimental runs.

Data	Dim	DFT	DWT	PAA	SVD	Plain
						Euclidean
TS01	5	0.89	0.76	0.92	0.91	0.92
	10	0.93	0.95	0.93	0.91	
TS05	5	0.93	0.67	0.94	0.94	0.99
	10	0.98	0.67	0.94	0.97	

Table 1 Clustering results

The results in the table clearly demonstrate the strong discriminative strength of the time series. The plain Euclidean achieves the best result since it takes into account all the details about the patterns. Observe though that with the dimensionality reduction techniques of time series, even after the dimensionality is reduced to as low as 5 or 10, the accuracy can remain above 90%. The relatively poor performance of DWT on TS05 is probably caused by the insufficiency of coefficients.

Here, we compare the proposed method with a static partitioning approach where each dimension is spilt into l equal-length bins, and we get a l*l*l hyper-rectangle for each 3D image. The mean values of the sub-regions are used as representative attributes. Using the Euclidean as the distance measure, we performed clustering experiments. As shown in Table 2, even with a dimensionality of 64 (when l=4), the resulting accuracy (0.61) is much worse than that of the proposed method.

Table 2 Clustering results for static partitioning method

l	Dim	Clustering Accuracy
2	8	0.56
3	27	0.58
4	64	0.61

3.2.2 K Nearest-neighbor (KNN) retrieval experiments

K nearest-neighbor retrieval is another important application: given a query subject, the k most similar subjects are to be retrieved. The more subjects within the same class as the query are found, the better the retrieval results. Precision and recall are a pair of metrics used to evaluate the effectiveness of the retrieval: precision is the number of subjects with the same class label divided by k, and recall is the number of subjects with the same class label divided by the total number of subjects in the same class. It is desirable that the precision stays at a high level when the recall ratio increases from 0 to 1.

To get a stable result, we applied 9-fold cross validation in our experiments: each time one control and one patient are taken as a query and their k nearest neighbors are retrieved. The averaged precision and recall ratios are reported in Figure 3 for the two time series datasets.



Figure 3. *k* nearest-neighbor retrieval results on (a) Dataset TS01 (b) Dataset TS05

For the plain Euclidean method, the accuracy of retrieval is nearly always perfect, with 100% precision. Even after the dimensionality is reduced to as low as 10, we can still achieve very good precision (between 80-100%) with the exception of DWT on TS05. Again, all these results demonstrate the discriminative significance of the patterns we extracted using the Hilbert space-filling curve.

4. Conclusions

The main contribution of this paper is the application of techniques from the time series domain to data mining and analysis of spatial patterns in 3D images. By mapping the 3D space to 1D with the traversal of the Hilbert spacefilling curve and applying statistical tests of significance in the linear domain, we can detect discriminative patterns, which can be represented as time series. We propose the employment of analysis methods in the time series domain to examine similarities among the original 3D images. As a case study, we analyzed an fMRI dataset. Based on similarity measures in time series, we performed further applications, such as clustering and nearestneighbor retrieval resulting in high efficiency and accuracy.

5. ACKNOWLEDGEMENT

The authors would like to thank A. Saykin for providing the fMRI data set and clinical expertise. This work was supported in part by the NSF under grants IIS-0083423 and IIS-0237921, by the NIH under grant R01-MH068066-01, and by the Pennsylvania Department of Health. The funding parties specifically disclaim responsibility for any analyses, interpretations and conclusions.

6. REFERENCES

- V. Megalooikonomou, J. Ford, L. Shen, F. Makedon, F. Saykin, "Data mining in brain imaging," Statistical Methods in Medical Research, vol. 9, no. 4, pp. 359-394, 2000.
- [2] D. Kontos, V. Megalooikonomou, N. Ghubade, C. Faloutsos, "Detecting discriminative functional MRI activation patterns using space filling curves", in Proceedings of the 25th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), Cancun, Mexico, pp. 963-967, Sep. 2003
- [3] D. Hilbert. "Uber die stetige Abbildung einer Linie auf Flachenstuck," Math. Annln., 38:459--460, 1891.
- [4] B. Moon, H.V. Jagadish, C. Faloutsos, J.H. Saltz." Analysis of the Clustering Properties of the Hilbert Space-Filling Curve," IEEE Transactions on Knowledge and Data Engineering, 13(1):124--141, Feb. 2001.
- [5] K.J. Friston, A.P. Holmes, K.J. Worsley, J.P. Poline, CD. Frith, R.S.J. Frackowiak, "Statistical parametric maps in functional imaging: a general linear approach," Human Brain Mapping, 2: 189–210, 1995
- [6] Faloutsos, C., Ranganathan, M. & Manolopoulos, Y. Fast subsequence matching in time-series databases. Proceedings of the ACM SIGMOD Int'l Conference on Management of Data. Minneapolis, MN, May 25-27. pp. 419-429.,1994
- [7] Keogh, E., Chakrabarti, K., Pazzani, M. & Mehrotra, S. Dimensionality Reduction for Fast Similarity Search in Large Time Series Databases. Journal of Knowledge and Information Systems, 2000.
- [8] Agrawal, R., Faloutsos, C., & Swami, A. Efficient similarity search in sequence databases. Proc. of the 4thConference on Foundations of Data Organization and Algorithms, 1993.
- [9] Magnus Lie Hetland. A survey of recent methods for efficient retrieval of similar time sequences. To appear in Data Mining in Time Series Databases. World Scientific, 2004.
- [10] Yi-Leh Wu, Divyakant Agrawal, Amr El Abbadi: A Comparison of DFT and DWT based Similarity Search in Time-Series Databases. CIKM 2000: 488-495
- [11] Keogh, E., Chakrabarti, K., Pazzani, M. & Mehrotra, S. Locally adaptive dimensionality reduction for indexing large time series databases. In proceedings of ACM SIGMOD Conference on Management of Data. Santa Barbara, CA, May 21-24. pp 151-162, 2001.
- [12] Gavrilov, M., Anguelov, D., Indyk, P. & Motwani, R. Mining the stock market: Which measure is best? Proceedings of the International Conference on Data Mining and Knowledge Discovery (KDD '00), pp. 487-496,2000.