

Discovering Multivariate Motifs using Subsequence Density Estimation and Greedy Mixture Learning

David Minnen and Charles L. Isbell and Irfan Essa and Thad Starner

Georgia Institute of Technology
College of Computing / School of Interactive Computing
Atlanta, GA 30332-0760 USA
{dminn,isbell,irfan,thad}@cc.gatech.edu

Abstract

The problem of locating motifs in real-valued, multivariate time series data involves the discovery of sets of recurring patterns embedded in the time series. Each set is composed of several non-overlapping subsequences and constitutes a motif because all of the included subsequences are similar. The ability to automatically discover such motifs allows intelligent systems to form endogenously meaningful representations of their environment through unsupervised sensor analysis. In this paper, we formulate a unifying view of motif discovery as a problem of locating regions of high density in the space of all time series subsequences. Our approach is efficient (sub-quadratic in the length of the data), requires fewer user-specified parameters than previous methods, and naturally allows variable length motif occurrences and non-linear temporal warping. We evaluate the performance of our approach using four data sets from different domains including on-body inertial sensors and speech.

Introduction

Our goal is to develop intelligent systems that understand their environment by autonomously learning new concepts from their perceptions. In this paper, we address one form of this problem where the concepts correspond to recurring patterns in the sensory data captured by the intelligent agent. Such recurring patterns are often referred to as perceptual primitives or *motifs* and correspond to sets of similar subsequences in the time-varying sensor data. For example, a motif discovery system could find unknown words or phonemes in speech data, learn common gestures in video of sign language, or allow a mobile robot to learn endogenously meaningful representations that help inform its actions.

Because intelligent systems rarely rely on a single sensor or only on discrete readings, our work focuses on the discovery of motifs in real-valued, multivariate time series. Even in situations when a single sensing device is used, it is often the case that multiple, simultaneous values are sensed (as in the many pixels that make up each frame of a video) or multiple features are extracted from a single reading (such as Mel frequency cepstral coefficients (MFCCs) extracted from a sound wave).

Unsupervised motif discovery is difficult due to the severe lack of information available *a priori*, coupled with the large

amount of data typically collected. Generally, the number of motifs, the corresponding number of occurrences, the duration of the occurrences, the shape or other characterizing statistics, and the location of the motifs in the time series data are all unknown. Discovery in time series data (compared to other, non-temporal sequential data) is further complicated by non-linear temporal warping that can mask subsequence similarity. Finally, because even small data sets may have hundreds of thousands of sensor readings, sub-quadratic computational complexity is necessary for practical applications.

In this paper, we frame motif discovery as the problem of locating regions of high density in the space of all time series subsequences. Such regions concisely capture the idea of a motif by combining similarity and frequency in a data-relative measure (see Figure 2. For a particular data set, a region of high density must have many occurrences in a small space. Casting motif discovery in terms of density estimation is beneficial because the imprecise notions of “small space” (equivalent to “high similarity”) and “many occurrences” are naturally made precise relative to the data itself.

Our approach to motif discovery leads to several contributions:

1. We provide a simple method for efficiently locating high density regions in the space of time series subsequences that avoids problems caused by trivial matches.
2. We detect motif occurrences using a form of greedy mixture learning that allows variable-length occurrences, guarantees a globally optimal fit to the time series data, and provides a natural measure of motif quality.
3. Compared to previous methods, our approach reduces the number of user-specified parameters and is less sensitive to the precise value of the remaining parameters.

Subsequence Density Estimation

Adopting a density estimation framework for motif discovery provides a principled foundation for characterizing motifs but still leaves several issues unresolved. One must still deal with the problem of trivial matches, bandwidth selection, locating density modes, and bounding high density re-

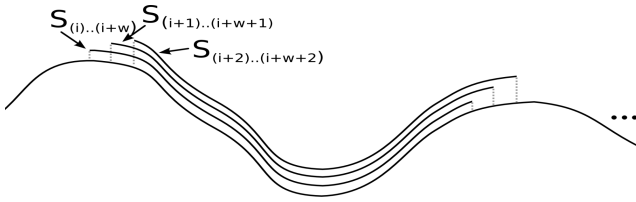


Figure 1: Three consecutive subsequences of length w from a 1D time series. The three subsequences are all very similar but are considered *trivial matches* since they overlap.

gions. We will address each of these problems in turn and present solutions that form the foundation of our motif discovery algorithm.

The Problem of Trivial Matches

When searching for motifs, we are interested in sets of time series subsequences that are highly similar. A trivial match occurs when a pair of subsequences are similar because they overlap and thus share many sensor readings (Keogh, Lin, & Truppel 2003). As an example of an extreme case, consider a time series S with subsequences $S^A = S_{i..i+w}$ and $S^B = S_{(i+1)..(i+w+1)}$ (see Figure 1). In smoothly varying time series, S^A and S^B will almost always be very similar, but this similarity does not imply a motif, nor does it imply a region of high density. To avoid such false detections, we require our motifs to be made up of non-overlapping subsequences and similarly restrict density estimates to only consider non-overlapping subsequences.

Bandwidth Selection for Subsequence Density Estimation

Typical methods for density estimation require specification of a bandwidth parameter that defines the size of the region of influence during the estimation process. Histograms, for example, require a bin size while kernel density estimates include a parameter that specifies the scale of the kernel. Many methods have been developed for computing a reasonable global bandwidth that attempt to balance bias and variance. These methods include heuristic formulae and optimization criteria such as minimization of the estimated mean integrated squared error (Silverman 1986). More robust methods provide for variable bandwidth estimation that automatically adapts the bandwidth size to match the local structure of the data (Sain 1999). Our approach uses an adaptive bandwidth based on a simple version of the balloon estimator first proposed by Loftsgaarden and Quesenberry (1965). In this method, the density is estimated directly from the distance to the k^{th} nearest neighbor.

Locating Density Modes

One of the most efficient and robust methods for locating density modes (local maxima) is the mean shift procedure (Comaniciu & Meer 2002). This algorithm iteratively computes a vector aligned with the estimated local gradient and then updates the mode estimate. Mean shift was used by Denton (2005) to find modes in fixed-length, univariate

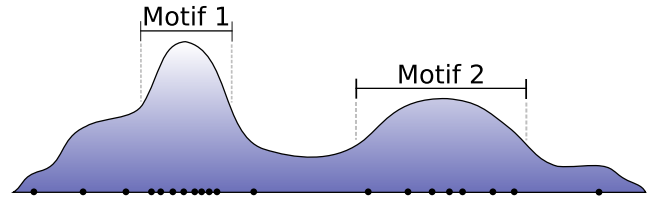


Figure 2: Subsequence density landscape in 1D. Each dot represents a different subsequence. Note that the two motifs coincide with regions of high density but that the motifs have different widths and different densities.

time series subsequences using the Euclidean distance metric. Our approach uses the dynamic time warping distance measure. This measure allows for non-linear temporal warping and has been shown to be much more robust for measuring the similarity of time series (Keogh & Ratanamahatana 2005).

In our approach, the subsequences located near each density mode are used to estimate the parameters of a hidden Markov model (HMM) that represents the corresponding motif. Thus a precise estimate for the location of the mode is not required so long as the subsequences near the mode are accurately detected. Given this relaxed requirement, we chose not to use mean shift and instead efficiently locate high density regions by first finding the k -nearest non-overlapping neighbors of each subsequence, then estimating the local density from the distance to the k^{th} neighbor, and finally selecting those subsequences with higher density than all of their neighbors. Although a naive implementation of k -nearest neighbor search requires $O(n^2)$ distance calculations, numerous methods have been proposed for building spatial indexes that reduce this to $O(n \log n)$ and dual-tree and approximate methods can often provide even better performance (Keogh & Ratanamahatana 2005; Liu 2006; Gray & Moore 2003).

Bounding High Density Regions

While the procedure presented above will identify candidate motifs seeds, it will neither detect all of the occurrences of each motif nor will it reliably select an accurate set of motifs. Incomplete motifs arise because there may be motif occurrences beyond the local maxima and its k -nearest neighbors. Many methods for addressing this deficiency exist. For instance, Chiu, Keogh, and Lonardi (2003) as well as Tanaka and Uehara (2005) utilize a user-specified radius to define a local, spherical neighborhood centered on each motif seed and then identify motif occurrences as all subsequences that lie within the hypersphere. Minnen *et al.* (2007) adopt a similar strategy but their approach uses a heuristic to automatically estimate the neighborhood radius from the data. Denton (2005) proposes a density-based subsequence clustering algorithm that estimates a density threshold from the subsequence data and an assumed random-walk noise model.

These approaches seek to directly bound the motif neighborhood either in terms of distance in subsequence space or in terms of density. We contend that these attempts

have shortcomings since they rely on restrictive assumptions (*e.g.*, spherical neighborhoods or random-walk noise) and require either accurate parameter estimation from noisy, under-sampled data or user specification. In contrast, our approach avoids direct estimation of the motif neighborhood size and instead lets the motif models compete to explain the time series data. This shift is accomplished by modeling each motif as a HMM and optimally fitting the set of HMMs to the time series data using a generalized Viterbi algorithm adapted from the speech recognition community (Young, Russell, & Thornton 1989).

The benefits of using a HMM-based continuous recognition system for motif occurrence detection are numerous. First, as noted above, we no longer need to directly estimate the motif neighborhood size. Second, the motif models are free to shrink or stretch to account for non-linear warping and variable length motif occurrences in the data, whereas the existing approaches are restricted to fixed-length subsequences. Third, the models provide a measure of goodness-of-fit in the form of the total data log-likelihood and the likelihood per motif. As explained in the following section, this evaluation metric will prove important for motif ranking and selection. Fourth, the continuous recognition algorithm is globally optimal in the sense that it finds the state sequence that maximizes the data log-likelihood. Fifth, this fitting procedure actually addresses both issues described in the beginning of this section: it locates all of the motif occurrences given the models and also simplifies the process of pruning redundant or spurious motif seeds. This pruning is possible since redundant or spurious motifs will score poorly due to the relatively small improvement in overall data log-likelihood that they provide.

Greedy Mixture Learning for Motif Selection

Algorithm 1 gives an overview of our density-based motif discovery algorithm. The previous section describes steps 1 through 4, which locate an over-complete set of candidate motif seeds by identifying subsequences located near high density regions. In order to select the correct motifs from this set and to find additional motif occurrences, our approach adopts a greedy mixture learning framework. This formulation was previously adopted by Blekas *et al.* (2003) for motif discovery in discrete, univariate sequences.

In traditional applications, the mixture components jointly explain the data by sharing responsibility for each data point (*i.e.*, $p(x) = \sum_{i=1}^N w_i p(x|\theta_i)$, where θ_i represents the parameters of the i^{th} component and w_i is the weight given to that component, constrained by $\sum_{i=1}^N w_i = 1$). The learning problem is to estimate the parameters, θ_i , the weights, w_i , and the number of components, N , that maximize the total data likelihood. In our context, however, motif occurrences do not overlap temporally, and so only one mixture component can be used to explain each frame of data and the set of components must jointly explain the data in a temporally exclusive manner. The pattern recognition community has developed efficient algorithms for solving such problems, principally for the purpose of continuous speech recognition (Young, Russell, & Thornton 1989). In

Algorithm 1 Density-based Motif Discovery

Input: Time series data (S), subsequence length (w), number of nearest neighbors to use (k), distance measure

Output: Set of discovered motifs including motif models and occurrence locations

1. Collect all subsequences, S_i of length w from the input data S
 2. Locate the k -nearest neighbors for each subsequence: $knn(S_i) = S_{i,1..k}$
 3. Estimate the density for each subsequence: $den(S_i) \propto 1/dist(S_i, S_{i,k})$
 4. Identify local maxima according to density: $maxima(S_i) = S_i : \forall S_{i,j} den(S_i) > den(S_{i,j})$
 5. Initialize the set of motif HMMs with a single background model: $H = \{bg\}$
 6. For each motif seed in $maxima(S_i)$:
 - (a) Construct $seed_i$, a HMM learned from the i^{th} density maxima and its k -nearest neighbors
 - (b) Fit the existing models plus the seed model ($H \cup seed_i$) to the time series data
 7. Greedily select the best motif seed: $m = \underset{i}{arg\ max} \log p(S|H \cup seed_i)$
 8. Test for stopping criteria for $H \cup seed_m$; if test fails set $H = H \cup seed_m$ and goto Step 6
 9. Re-estimate motif models in H and return $H/\{bg\}$
-

typical speech systems, each word is modeled by a HMM (commonly this is a composite model built from HMMs representing the relevant phonemes) and then each utterance is recognized by finding the mapping from word HMMs to the speech signal that maximizes the log-likelihood of the utterance (HTK 2007; Rabiner & Juang 1993). In our algorithm, HMMs constructed from motif seeds (*i.e.*, a local maxima in density space along with its k -nearest neighbors) take the place of the word models.

Although traditional parameter estimation methods for HMMs, such as the Baum-Welch algorithm, typically fail when applied to so few training examples, a simple construction algorithm is sufficient to capture the characteristics of each motif. Deficiencies in the resulting model are countered by the competitive continuous recognition framework and by a round of full parameter re-estimation using all identified motif occurrences. The HMM construction algorithm builds a model with a left-right topology and with the number of states equal to half the user-specified subsequence length (see Figure 3). Each training sequence is divided into equal-length, overlapping segments corresponding to each state. Each state has a Gaussian observation distribution with parameters estimated from its associated segment.

In order to select the best motif from the set of candidates, we directly optimize the conditional data likelihood by com-

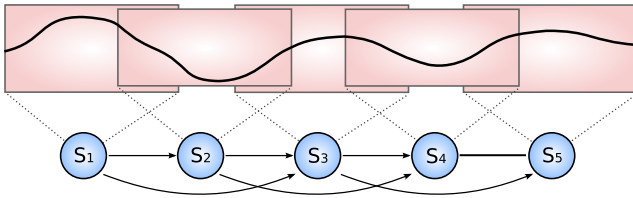


Figure 3: HMM construction: each sequence in the training set is divided into equal-size, overlapping segments from which the parameters of a Gaussian distribution are estimated for each state.

puting the optimal mapping for each seed and then selecting the one that results in the highest likelihood (Algorithm 1: Steps 5 - 7). This procedure is equivalent to ranking motif seeds by information gain, which is calculated as the change in data log-likelihood when the motif candidate is included in the set of discovered motifs. Thus, information gain is used as a contextual measure of motif quality that combines occurrence similarity and frequency. Viewed this way, we see how information gain and density are complementary. Information gain provides a global measure of motif quality relative to the data and other motif models, while density provides an efficient local measure useful for focusing computational resources on promising regions of subsequence space.

The final component of our approach is responsible for detecting when to stop adding motifs to the mixture model (Step 8). Previous methods used simple stopping criteria that rely on the restrictive assumptions or additional information that they require. For instance, the methods of Chiu *et al.* (2003), Minnen *et al.* (2007), and Tanaka and Uehara (2005) search for additional motifs until no pair of subsequences lie within the same neighborhood. Denton’s method (2005) continues searching until no subsequence lies in a region with density above a threshold estimated based on an assumed random-walk noise model.

In contrast, our approach uses the density estimate of each motif, computed after all occurrences have been detected, to determine when to stop. The algorithm searches for a local minima in the smoothed derivative of the motif densities, which corresponds to an inflection point in the curve representing the motif densities. This inflection point marks the transition from well-supported motifs (high density) to spurious motifs (low density). Although the usefulness of this heuristic method is demonstrated through the empirical results presented in the next section, developing a more well-grounded approach for estimating the correct number of motifs is a major part of our ongoing research.

Experimental Results

We have evaluated our approach to multivariate motif discovery using four data sets. The evaluation compares the discovered motifs to manually labeled patterns known to exist in the data and then computes the accuracy of the match by considering the number of correct detections, missed occurrences, false detections, and incorrect identifications. Note that this is a difficult error metric for continuous recog-

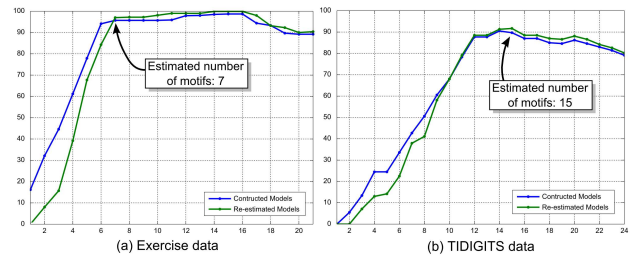


Figure 4: Graph of accuracy vs. number of discovered motifs for the exercise and TIDIGITS data sets. Our algorithm automatically selected seven motifs for the exercise data (*truth*: six + background) and 15 for the TIDIGITS data (*truth*: 11 + background).

nition systems and can actually give negative rates if a system reports too many false detections.

The first data set was captured during an exercise regime composed of six different dumbbell exercises. An Xsens MT9 inertial motion sensor was attached to the subject’s wrist and readings from a three-axis accelerometer and gyroscope were recorded at 100Hz. In total, roughly 27.5 minutes of data was captured across 32 sequences. For the experiment, the data was resampled at 12.5Hz leading to 20,711 frames. The data contains six different exercises and 864 total repetitions (roughly 144 occurrences of each exercise). Each frame is composed of the raw three-axis accelerometer and gyroscope readings leading to a six-dimensional feature vector.

The second evaluation uses the publicly available TIDIGITS data set (Leonard & Doddington 1993). We evaluate our method on a subset of the available data consisting of 77 phrases composed of spoken digits. The data contains 11 classes (zero through nine plus “oh”) and 253 total digit utterances (roughly 23 occurrences of each digit). We extract MFCCs over a range from zero to 4kHz using a frame period of 10ms and a window size of 25ms. This feature extraction process leads to a final data set containing 11,917 frames (see Figure 6).

Figure 4 shows the accuracy rate of our algorithm for the exercise and TIDIGITS data sets for different numbers of discovered motifs. In both cases, the algorithm identified all of the real motifs as well as additional motifs that correspond to recurring background patterns. Not surprisingly, the first unexpected motif learned from both data sets represents silence, either literally in the case of the TIDIGITS data or as a period of very little motion in the case of the exercise data.

Although the graphs show performance measured across a range of motifs, our approach automatically estimates the correct number as previously described. For the exercise data, the algorithm estimated seven motifs, leading to an accuracy of 95.7% initially and 97.0% after parameter re-estimation. This is significantly better than earlier methods. For comparison, the best parameter settings for Minnen *et al.*’s approach (2007) resulted in 12 motifs with an accuracy of 92.8%, but locating this parameter setting required considerable search over the various parameter settings followed by manual selection based on a comparison to the

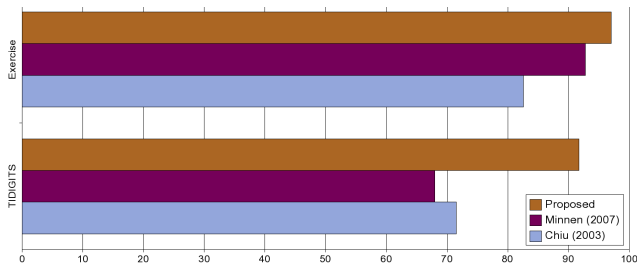


Figure 5: Comparison of motif accuracy on the exercise and TIDIGITS data sets.

known motifs. Our method, in contrast, only required a single run. Similarly, the best parameter setting for Chiu *et al.*'s method (2003) leads to an accuracy of 82.6% and required an even more extensive manual search than Minnen's algorithm.

On the spoken digits data, our algorithm estimated 15 motifs corresponding to an accuracy of 89.7% initially and 91.7% after model re-estimation. The best parameter setting led Minnen *et al.*'s method to find 11 motifs with an accuracy of 68.0%, while Chiu *et al.*'s algorithm also found 11 motifs with an overall accuracy of 71.5%. See Figure 5 for a summary of these results.

The final two data sets are taken from the UCR Time Series Data Mining Archive (Keogh & Folias 2002) and have been used to demonstrate the generality of previous multivariate motif discovery algorithms (Tanaka, Iwamoto, & Uehara 2005; Minnen *et al.* 2007). The data represents shuttle telemetry (6D, 1000 frames) and ECG readings (8D, 2500 frames). Both data sets contain a single motif, which our algorithm correctly identifies. In addition, our algorithm detects the motif occurrences with 100% accuracy, including one partial occurrence in the ECG data that the previous methods failed to detect.

Related Work

Researchers from bioinformatics, data mining, and artificial intelligence have looked at variations of the motif discovery problem. Within the bioinformatics community, researchers have focused on discrete, univariate sequences such as those that represent nucleotide sequences. For instance, MEME uses expectation maximization (EM) to estimate the parameters of a probabilistic model for each motif (Bailey & Elkan 1994). GEMODA unifies several earlier methods but requires computing pair-wise distances between subsequences leading to a quadratic expected run time (Jensen *et al.* 2006). Finally, Blekas *et al.* (2003) adapted a method for spatial greedy mixture learning to sequential motif discovery, which inspired our use of continuous recognition and information gain.

Data mining researchers have developed several methods for motif discovery in real-valued, univariate data. Chiu *et al.* (2003) use a local discretization procedure and random projections to find candidate motifs in noisy data. Denton's (2005) approach is more closely related to our method as it avoids discretization and frames subsequence

clustering in terms of kernel density estimation. Her method differs from ours in many ways including the use of the Euclidean distance metric and the assumption of a random-walk noise model and univariate data.

Minnen *et al.* extend Chiu's approach by supporting multivariate time series and automatically estimating the neighborhood size of each motif. Tanaka *et al.* (2005) also extend Chiu's works to support multidimensional data and variable-length motifs. Their method uses principal component analysis to project multivariate data down to a single dimension and then applies a univariate discovery algorithm. They report promising results for several data sets, but it is straightforward to show that the first principal component does not generally capture sufficient information to learn from multivariate signals. Finally, Oates (2002) developed the PERUSE algorithm to find recurring patterns in multivariate sensor data collected by robots. PERUSE allowed a robot to detect perceptual patterns in its sensor data and found many recurring phrases when applied to speech. PERUSE is also one of the few algorithms that can handle non-uniformly sampled data, but it suffers from some computational drawbacks and stability issues when estimating motif models.

Discussion & Future Work

The strength of our approach stems from formulating motif discovery in terms of subsequence density estimation and greedy mixture learning. Both density and information gain provide principled measures of motif quality for which efficient algorithms have been developed. The major drawback of our approach is the need to specify a canonical subsequence length used to compute an initial density estimate. Although the continuous recognition system allows the final motif occurrences to vary in length, the initial restriction can lead to poor density estimates and increases the burden on users to specify an accurate value. To improve this situation, we are investigating the feasibility of searching over a range of subsequence lengths. Our current formulation should scale naturally to this more general problem since it already handles overlapping subsequences and redundant motif seeds.

By reinterpreting previous methods as subsequence density estimation, we can draw some interesting comparisons. For instance, the methods of Chiu *et al.* (2003) and Minnen *et al.* (2007) both search for pairs of similar subsequences using a very efficient discrete representation and random projection method. This can be seen as a rough density approximation based on a single nearest neighbor. Viewed this way, it makes sense to experiment with their approach as a way of finding candidate motif seeds for use with our greedy mixture learning framework.

Our primary goal for future research is to generalize the meaning of multivariate motifs. Currently, a motif must exist across all of the dimensions. For derived features such as MFCCs in speech, this definition is reasonable. For other scenarios such as mobile robotics and distributed sensor systems, a more realistic definition allows motifs to exist across only a subset of the dimensions. Our current research is focused on developing an efficient algorithm for discovering such "sub-dimensional" motifs.

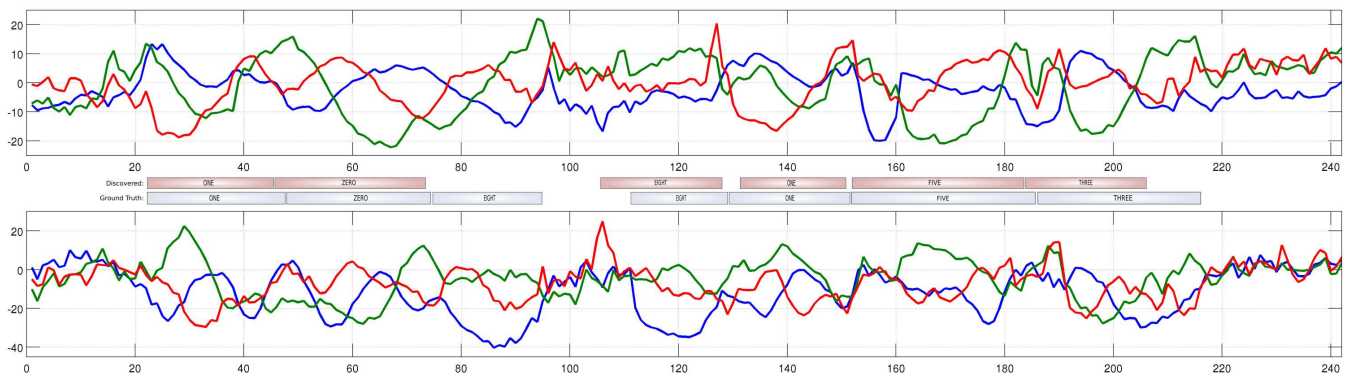


Figure 6: First six MFCCs extracted from the utterance “one zero eight eight one five three.” The center bars show the discovered motifs (top) and ground truth labels (bottom). The discovered occurrences are quite accurate, though the first “eight” was not detected.

Conclusion

Equating time series motifs with regions of high density in the space of all subsequences provides a principled framework in which to design a discovery algorithm. We use density estimation to identify promising candidate motifs and rely on a continuous recognition system within a greedy mixture learning framework to select motifs and detect additional occurrences. Evaluation of this approach on speech and on-body sensor data demonstrate that it can discover motifs with higher accuracy than previous methods while requiring fewer user-specified parameters.

References

- Bailey, T., and Elkan, C. 1994. Fitting a mixture model by expectation maximization to discover motifs in biopolymers. In *Second International Conference on Intelligent Systems for Molecular Biology*, 28–36. AAAI Press.
- Blekas, K.; Fotiadis, D.; and Likas, A. 2003. Greedy mixture learning for multiple motif discovery in biological sequences. *Bioinformatics* 19(5).
- Chiu, B.; Keogh, E.; and Lonardi, S. 2003. Probabilistic discovery of time series motifs. In *Conf. on Knowledge Discovery in Data*, 493–498.
- Comaniciu, D., and Meer, P. 2002. Mean shift: A robust approach toward feature space analysis. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 24:603–619.
- Denton, A. 2005. Kernel-density-based clustering of time series subsequences using a continuous random-walk noise model. In *Proceedings of the Fifth IEEE International Conference on Data Mining*.
- Gray, A. G., and Moore, A. W. 2003. Very fast multivariate kernel density estimation via computational geometry. In *Proceedings of the ASA Joint Statistical Meeting*.
- HTK. 2007. HTK Speech Recognition Toolkit. Machine Intelligence Laboratory, Cambridge University. <http://htk.eng.cam.ac.uk>.
- Jensen, K.; Styczynski, M. P.; Rigoutsos, I.; and Stephanopoulos, G. 2006. A generic motif discovery algorithm for sequential data. *Bioinformatics* 22(1):21–28.
- Keogh, E., and Folias, T. 2002. UCR time series data mining archive.
- Keogh, E., and Ratanamahatana, C. A. 2005. Exact indexing of dynamic time warping. *Knowledge and Information Systems* 7(3):358–386.
- Keogh, E.; Lin, J.; and Truppel, W. 2003. Clustering of time series subsequences is meaningless: Implications for past and future research. In *ICDM*, 115–122.
- Leonard, R. G., and Doddington, G. 1993. TIDIGITS Linguistic Data Consortium, Philadelphia.
- Liu, T. 2006. *Fast Nonparametric Machine Learning Algorithms for High-dimensional Massive Data and Applications*. Ph.D. Dissertation, Carnegie Mellon University.
- Loftsgaarden, D., and Quesenberry, C. 1965. A nonparametric estimate of a multivariate density function. *The Annals of Mathematical Statistics* 36:1049–1051.
- Minnen, D.; Starner, T.; Essa, I.; and Isbell, C. 2007. Improving activity discovery with automatic neighborhood estimation. In *International Joint Conference on Artificial Intelligence*.
- Oates, T. 2002. PERUSE: An unsupervised algorithm for finding recurring patterns in time series. In *Int. Conf. on Data Mining*, 330–337.
- Rabiner, L., and Juang, B.-H. 1993. *Fundamentals of Speech Recognition*. Signal Processing Series. Prentice Hall.
- Sain, S. 1999. Multivariate locally adaptive density estimation. Technical report, Department of Statistical Science, Southern Methodist University.
- Silverman, B. 1986. *Density Estimation*. London: Chapman and Hall.
- Tanaka, Y.; Iwamoto, K.; and Uehara, K. 2005. Discovery of time-series motif from multi-dimensional data based on mdl principle. *Machine Learning* 58(2-3):269–300.
- Young, S.; Russell, N.; and Thornton, J. 1989. Token passing: a simple conceptual model for connected speech recognition systems. Technical Report 38, Cambridge University.