

# Semi-Markov Modeling-Clustering of Human Sleep with Efficient Initialization and Stopping

Chiying Wang<sup>1</sup>, Sergio A. Alvarez<sup>2</sup>, Carolina Ruiz<sup>1</sup> and Majaz Moonis<sup>3</sup>

<sup>1</sup>Department of Computer Science, Worcester Polytechnic Institute, Worcester, MA 01609 U.S.A.

<sup>2</sup>Department of Computer Science, Boston College, Chestnut Hill, MA 02467 U.S.A.

<sup>3</sup>Department of Neurology, University of Massachusetts Medical School, Worcester, MA 01605 U.S.A.

**Keywords:** Time Series, Dynamic Time Warping, Data Mining: Clustering, Modeling, Markov, Sleep.

**Abstract:** Collective Dynamical Modeling-Clustering (CDMC) is an algorithmic framework for time series dynamical modeling and clustering using probabilistic state-transition models. In this paper, an efficient initialization technique based on Itakura slope-constrained Dynamic Time Warping is applied to CDMC. Semi-Markov chains are used as the dynamical models. Experimental evaluation demonstrates the effectiveness of the proposed approach in providing more realistic dynamical modeling of sleep stage dynamics than Markov models, with improved clustering quality and convergence speed as compared with pseudorandom initialization.

1 INTRODUCTION

The dynamics of the sleep-wake cycle present species-specific patterns with some underlying similarities across mammalian species (Phillips et al., 2010). In humans, a more detailed description of sleep is available in terms of the sleep stages determined via polysomnography, as described in the Rechtschaffen-Kales (Rechtschaffen and Kales, 1968) and AASM (Iber et al., 2007) staging standards. An example of an all-night sleep stage sequence is shown in Fig. 1. Sleep stage transitions and sleep stage durations provide essential indicators in describing the relationship between sleep and health (Kishi et al., 2008), (Hernandez et al., 2009), (Chervin et al., 2009), and (Bianchi et al., 2010).

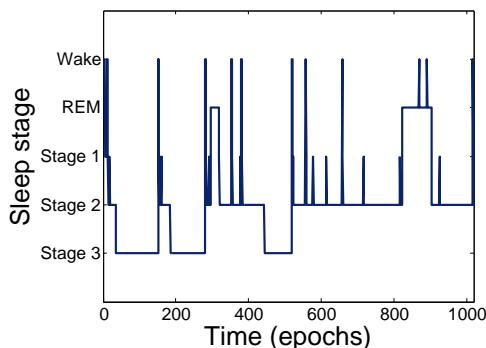


Figure 1: Sample hypnogram

**Scarcity of Dynamical Events in Sleep Data.** We are interested in the discovery of dynamical patterns in sleep stage sequences by means of automated algorithms. Sleep presents a particular challenge to such algorithms, in that key dynamical events such as stage transitions occur very sparsely within a hypnogram, making it difficult to extract reliable dynamical information from a single night of sleep for a given individual (Bianchi et al., 2010).

The collective dynamical modeling-clustering (CDMC) algorithm (Alvarez and Ruiz, 2013) addresses the problem of scarcity of dynamical events by pooling sleep data across multiple individuals, simultaneously partitioning the collection of subjects by sleep-dynamical similarity. CDMC reduces the model variance by selectively aggregating instances through clustering. This reduction in variance is accomplished in CDMC without the loss of detail that would result by simply aggregating data without regard for dynamical similarity. The result of CDMC is a collection of groups of hypnograms such that hypnograms within a given group are dynamically similar, while hypnograms in different groups are not.

**Scope of the Paper.** CDMC is a general algorithmic framework that leaves open several important choices. Two main choices that must be made when applying the CDMC framework are the clustering initialization technique and the dynamical model type. Stan-

dard choices are pseudorandom clustering initialization, and Markov dynamical models. These are the choices used in (Alvarez and Ruiz, 2013). Markov models have been widely used in previous applications to sleep (Zung et al., 1965) and (Kim et al., 2009).

The present paper focuses on two main elements:

1. The use of Dynamic Time Warping (DTW) for clustering initialization in CDMC. DTW has previously been applied to initialization of a different clustering technique based on Hidden Markov Models (Oates et al., 1999). We show that DTW initialization significantly improves CDMC convergence as compared with pseudorandom initialization.
2. The use of semi-Markov dynamical models in CDMC in order to better describe the dynamical characteristics of human sleep as compared with Markov models.

Experimental evaluation confirms the advantages of the approach proposed in the paper.

The rest of the paper is organized as follows. Section 2 briefly reviews the CDMC algorithm, presents the application of dynamic time warping to clustering initialization in CDMC, and describes the semi-Markov dynamical model version of CDMC. Section 3 describes the synthetic and actual human sleep datasets used for evaluation, together with the experimental protocol. Section 4 presents experimental results and analysis of the proposed approach. Section 5 describes conclusions and ideas for future work.

## 2 PROPOSED APPROACH

This section describes the Collective Dynamical Modeling-Clustering (CDMC) framework that constitutes the foundation for this work (Algorithm 1), the semi-Markov chains to be used as the dynamical models in CDMC, the Dynamic Time Warping-driven clustering initialization approach for CDMC (Algorithm 2), and the clustering similarity metrics to be used for the CDMC stopping criteria.

### 2.1 Collective Dynamical Modeling-Clustering (CDMC)

CDMC is a general algorithmic strategy for simultaneous clustering and dynamical modeling of sequence data (Alvarez and Ruiz, 2013) that is shown in pseudocode in Algorithm 1. CDMC simultaneously groups data instances by dynamical similarity

and induces a dynamical model of each group. The following is an outline of the CDMC procedure:

- An initial grouping of instances  $x_1, \dots, x_n$  into  $k$  clusters is provided.
- The grouping is iteratively refined by repeating the following steps until the similarity of two successive clusterings (e.g.,  $c$  and  $c_{old}$ ) reaches a pre-determined similarity threshold  $\text{minSim}$  (step 3):
  - A maximum likelihood dynamical model  $M_i$  is built from each cluster  $C_i$  (step 5).
  - Each instance  $x$  is assigned to the cluster  $C(x)$  for which the generative likelihood  $P(x | M_{C(x)})$  is maximized (step 6).
- A final clustering  $c$  of the dataset and a generative model  $M_i$  for each of the clusters are returned (step 7).

Note that the details of cluster initialization, dynamical model type, and similarity metric are left unspecified in the general version of CDMC. (Alvarez and Ruiz, 2013) includes an illustration that uses pseudorandom initialization, hidden Markov models, and the Rand index in these roles. The present paper shows that improved results can be obtained by using agglomerative distance-based clustering with dynamic time warping for initialization, and semi-Markov chains as the dynamical models.

### 2.2 Semi-Markov Dynamical Models for CDMC

Markov chains and the related models with partially observable state spaces, hidden Markov models, have been used to model sleep stage transitions in previous work (Zung et al., 1965), (Kim et al., 2009), (Alvarez and Ruiz, 2013). State transitions in a Markov model occur with a fixed probability in every cycle of a standard clock that is shared by all of the states. The probability of transitioning from state  $i$  to state  $j$  in a given clock cycle is a fixed probability value  $p_{i,j}$ . In particular, the probability of leaving state  $i$  in a given clock cycle of a Markov model is the fixed value  $p_{i,*} = \sum_{j \neq i} p_{i,j}$ . It follows that the probability of remaining in state  $i$  for precisely  $n$  clock cycles, leaving in the  $(n+1)$ st cycle, is  $(1 - p_{i,*})^n p_{i,*}$ . Thus, the duration of visits to a given state in a Markov model has a geometric (discrete exponential) probability distribution.

Experimental results (Kim et al., 2009) and (Chu-Shore et al., 2010) demonstrate that geometric distributions are a poor fit for actual stage bout duration

**Algorithm 1:** Collective Dynamical Modeling-Clustering (CDMC, (Alvarez and Ruiz, 2013)).

**Input:** An unlabeled time-series dataset  $D = \{x = (a_i(x))|i = 1, 2, \dots, n\}$ ; a positive integer,  $k$ , for the desired number of clusters; an initial guess  $c_0 : D \rightarrow \{1, \dots, k\}$  of the cluster label  $c_0(x)$  of each instance  $x \in D$ ; parameter values,  $s$ , specifying the desired configuration of the models (e.g., number of states); and a real number  $\text{minSim}$  between 0 and 1 for the minimum clustering similarity required for stopping.

**Output:** A set  $M_1, \dots, M_k$  of generative dynamical models (with configuration parameters  $s$ ), together with a cluster labeling  $c : D \rightarrow \{1, \dots, k\}$  that associates to each data instance,  $x$ , the index  $c(x)$  of a model  $M = M_{c(x)}$  for which the generative likelihood  $\prod_{x \in D} P(x|M_{c(x)})$  is as high as possible.

$\text{CDMC}(D, k, c_0, s, \text{minSim})$

1.  $c(x) = c_0(x)$  for all  $x$  in  $D$
2.  $c_{old}(x) = 0$  for all  $x \in D$
3. **while**  $\text{CLUSTERINGSIMILARITY}(c, c_{old}) < \text{minSim}$
4.      $c_{old} = c$
5.      $(M_1, \dots, M_k) = \text{LEARNMLPROTOTYPES}(D, k, c, s)$
6.      $c = \text{LEARNMLCLUSTERLABELS}(D, M_1, \dots, M_k)$
7. **return**  $M_1, \dots, M_k, c$

distributions of human sleep.<sup>1</sup> For example, the wake stage has a distribution of bout durations with a slowly decaying tail that is more similar to a power-law function than to a discrete exponential. Motivated by this fact, the present paper employs a semi-Markov chain model for the sequence of sleep stages, instead of a Markov model. In a semi-Markov model, the mechanism that determines the durations of state visits is independent of any model-wide clock. Each state  $i$  in a semi-Markov model of the type considered here has a specified visit duration distribution  $P_i(\tau)$ ; upon arriving in the given state, a random sample  $\tau$  is taken from this visit duration distribution; the model then remains in state  $i$  for the duration  $\tau$ , at the end of which time the next state is selected from among all states other than  $i$ , according to some Markov-type transition probability matrix with zeros along the diagonal. Other varieties of semi-Markov models also exist. For a general discussion, see (Yu, 2010).

Prior research (Wang et al., 2013) has established that individual sleep stage durations are well modeled by Weibull distributions. The present paper therefore uses semi-Markov dynamical models in which the state visit duration distributions belong to the Weibull family. A completely observable state space is used, with one state for each sleep stage considered. The resulting semi-Markov chains are used as the dynamical models within the CDMC framework discussed above in section 2.1. Standard Markov models within CDMC provide a benchmark for performance comparisons.

<sup>1</sup>A *stage bout* is a maximal uninterrupted period during which the stage of sleep remains the same.

## 2.3 Clustering Initialization and Stopping Criteria for CDMC

This section proposes using agglomerative clustering with a distance function based on Dynamic Time Warping as the clustering initialization technique for CDMC. This idea is based on prior work (Oates et al., 1999) that uses Dynamic Time Warping in conjunction with Hidden Markov Models to cluster time series data. The use of the modified Rand index as a similarity metric in the CDMC stopping criterion in order to account for the role of chance is also discussed.

### 2.3.1 Dynamic Time Warping (DTW)

Dynamic Time Warping (DTW) is a classical dynamic programming algorithm that provides an optimal alignment between two time series by nonlinearly warping their time dimensions (Berndt and Clifford, 1994). DTW has been extensively used in speech recognition, periodic movement capture, and so on (Sakoe and Chiba, 1978) and (Itakura, 1975). In this paper, DTW is used as a measure of similarity for unsupervised clustering of time series in section 2.3.2.

The standard dynamic programming approach to DTW for input sequences of length  $n$  implicitly considers all pairings of time indices in the two input sequences, leading to  $O(n^2)$  time complexity. Suboptimal constrained versions of DTW aim to reduce the time complexity by restricting the portion of the index space considered in the warping search. Itakura slope-constrained DTW (Itakura, 1975) imposes slope constraints on the warping path, thereby constraining the search for a warping path to a parallelogram in the

index space. The Sakoe-Chiba approach (Sakoe and Chiba, 1978) constrains the warping search to a band in the index space. An alternative hierarchical multiresolution approach has the potential to reduce the time complexity of DTW to  $O(n)$  (Salvador and Chan, 2007).

### 2.3.2 DTW Clustering Initialization (DTWC)

The core of the clustering initialization approach for CDMC that is proposed in the present paper is described in pseudocode in Algorithm 2. This approach performs agglomerative metric clustering using the distance function computed by DTW. The main steps of the proposed DTWC initialization for CDMC are:

- Place each instance  $x_1, \dots, x_n$  in its own cluster  $C_1, \dots, C_n$  (step 1-2)
- Repeat until there are only  $k$  clusters left (step 3)
  - Merge the closest clusters,  $C_i$  and  $C_j$ ; the distance measure between two instances (i.e.,  $x_s$  and  $x_t$ ) is defined by DTW; the distance measure between two clusters is the average distance of instances in these clusters (step 4-5)
- Return the final partition of the dataset  $D$  into  $k$  clusters (step 7)

**Constrained DTW-Driven Clustering (cDTWC).** A faster variant of DTWC, constrained DTWC (cDTWC), is obtained by using Itakura slope-constrained DTW instead of the full DTW when computing the distance metric in Algorithm 2. These two initialization techniques for CDMC will be compared with pseudorandom initialization in section 4.2.

### 2.3.3 Stopping Criteria

The Rand index (RI) (Rand, 1971) is a measure of agreement between two partitions of the same set, and therefore could be used in Algorithm 1 to measure the similarity of two clusterings on the same data. A drawback of this index is that it will produce a nonzero value for the comparison of two randomly constructed partitions. The adjusted Rand index (ARI) (Hubert and Arabie, 1985) is based on the Rand index but corrects for clustering agreements due to chance. Normalized mutual information (NMI) (Vinh et al., 2010) is a distinct way of evaluating clusters by the tradeoff between the number of clusters and qualities. In the present paper, these three metrics are compared as the basis for the stopping criterion in CDMC (Algorithm 1), using the resulting CDMC convergence time (number of modeling-

---

**Algorithm 2:** DTW-driven Clustering (DTWC).

---

**Input:** An unlabeled time series dataset  $D = \{x_1, x_2, \dots, x_n\}$ ; a positive integer,  $k$ , for the desired number of clusters; a predefined local distance measure  $d : \mathcal{X} \times \mathcal{X} \rightarrow \mathbb{R}_{\geq 0}$  where  $\mathcal{X}$  denotes the space in which the  $x_i$  take their values.

**Output:** A partition  $C$  of  $D$  into  $k$  clusters.

$DTWC(D, k, d)$

1.  $C_i = \{x_i\}$  for each  $x_i$  in  $D$
  2.  $m = n$
  3. **while**  $m > k$
  4.  $(i^*, j^*) = \arg \min_{i,j \in \{1, \dots, m\}} \bar{d}(C_i, C_j)$   
 $= \arg \min_{i,j \in \{1, \dots, m\}} \left\{ \frac{\sum_{l(x_s)=C_i, l(x_t)=C_j} DTW(x_s, x_t, d)}{|C_i| \cdot |C_j|} \mid x_s, x_t \in D \right\}$
  5. Merge  $C_{i^*}$  and  $C_{j^*}$  so that  $C = \{1, \dots, m-1\}$
  6.  $m = m - 1$
  7. **return**  $\{C_1, \dots, C_k\}$
- 

clustering iterations) and classification accuracy over labeled synthetic data for evaluation (see section 4).

## 3 EVALUATION

This section describes the synthetic and human sleep datasets, as well as the protocols used for experimental evaluation of the approach proposed in section 2.

### 3.1 Data Description

#### 3.1.1 Human Sleep Data

A collection of 200 fully anonymized human polysomnographic recordings was extracted from polysomnographic overnight sleep studies performed in the Sleep Clinic at Day Kimball Hospital in Putnam, Connecticut, USA, with approval of the respective Institutional Review Board for human subjects research. The subjects of the recordings had requested consultation due to sleep-related symptoms (e.g., sleepiness during daytime hours), and hence there is selection bias in the data. Each polysomnographic recording is split into 30-second epochs. Staging of each 30-second epoch into one of the standard sleep stages (wake, stage 1, stage 2, stage 3, and REM) is carried out by trained sleep technicians. Stages 1, 2, and 3 were subsequently grouped together into a single non-REM stage (NREM), resulting in a representation of human sleep in terms of three stages: Wake, NREM, and REM.

### 3.1.2 Semi-Markov Mixture Data

A synthetic dataset was also generated from two distinct semi-Markov chains, each with two states, but with different transition probability matrices and state duration statistics. The use of synthetic data provides precise control over the generative statistical parameters, whereas the generative parameters of human sleep data can only be estimated. Weibull state-duration distributions are used for each of the semi-Markov chains (see section 2.2).

**Semi-Markov Chain Parameter Values.** The following transition matrices and Weibull (shape, scale) parameter values (in that order) were used for the two semi-Markov chains. There are two Weibull distributions per chain, one for each state.

$$\begin{aligned} \text{chain 1 : } & \begin{pmatrix} 0.90 & 0.10 \\ 0.10 & 0.90 \end{pmatrix} \quad (3,5), \quad (2.5,4.5) \\ \text{chain 2 : } & \begin{pmatrix} 0.15 & 0.85 \\ 0.85 & 0.15 \end{pmatrix} \quad (3,4), \quad (2.5,3.5) \end{aligned}$$

**Generation of Synthetic Mixture Data.** In order to generate a sequence of a given length,  $L$ , from the mixture model, a random choice is first made between the two semi-Markov chains. A state sequence of the desired length is then generated by the randomly selected semi-Markov chain. The random chain selection and state sequence generation process continues until a desired total number of sequences,  $N$ , has been generated. See Algorithm 3.

In the present paper,  $N = 100$  sequences, each of length  $L = 100$ , were used in all trials involving synthetic semi-Markov data.

---

**Algorithm 3:** Generation of synthetic data.

---

**Input:** Positive integers  $N$  and  $L$ , semi-Markov chains  $S_1, \dots, S_k$ .

**Output:** A collection of  $N$  sequences, each of length  $L$ , generated by a mixture of the semi-Markov chains  $S_1, \dots, S_k$ .

`genData( $N, L, (S_1, \dots, S_k)$ )`

1.  $C = \{\}$
  2. **for**  $i = 1, \dots, N$
  3.      $j = \text{random choice among } \{1, \dots, k\}$
  4.      $x_i = \text{sequence of length } L \text{ generated by } S_j$
  5.      $C = C \cup \{x_i\}$
  6. **return**  $C$
- 

## 3.2 Experiment Description

Each experiment involved 100 trials. The human sleep dataset of section 3.1.1 was used in the comparative evaluation of semi-Markov and standard Markov dynamical models (section 4.1). The synthetic semi-Markov mixture dataset of section 3.1.2 was used in the evaluation of initialization techniques and stopping metrics. This allows the use of classification accuracy to evaluate performance (section 3.3). For the initialization experiments in section 4.2, the local cost measure was defined as that if a pair of elements in two sequences are same, the cost is 0, otherwise, 1. In section 4.3, the stopping metrics described in section 2.3.3: the standard Rand index (RI), the adjusted (chance-corrected) Rand index (ARI), and Normalized Mutual Information (NMI), were used respectively in CDMC, and the results were compared in terms of the resulting classification accuracy and number of iterations to convergence. The similarity threshold for stopping in CDMC (Algorithm 1) was set to 0.8. Wilcoxon-Mann-Whitney statistical hypothesis testing was used for comparison of medians (section 3.4). All experiments were performed in MATLAB® (*The MathWorks*, 2012).

## 3.3 Cluster Validity

Clustering quality in section 2.3 is evaluated by comparing the clustering results to known class labels for the synthetic data described in section 3.1.2. The class label of an instance is the Markov chain that generates that instance. The classification accuracy (fraction of instances that are classified correctly, relative to the set of all instances) is used as the evaluation metric. Higher classification accuracy of a clustering indicates a more meaningful data partition.

## 3.4 Statistical Significance

Non-normality of the observed accuracy and convergence time distributions was detected in some of the experiments, making the standard  $t$ -test for the difference of population means inappropriate. Comparison of population medians was therefore carried out by using a non-parametric Wilcoxon rank sum (Mann-Whitney) test at the  $p < 0.05$  significance level.

## 4 RESULTS

This section discusses the results obtained using CDMC with the proposed choices of semi-Markov

dynamical models, DTW-distance clustering initialization, and adjusted Rand index stopping criterion, as discussed in section 2. The evaluation protocol is discussed in section 3.

## 4.1 Semi-Markov Model

The human sleep data described in section 3.1.1 were clustered using CDMC (Algorithm 1) with  $k = 2$  clusters, for each of two dynamical model types: semi-Markov chains with Weibull state durations (section 2.2), and Markov chains. Three-state chains were used in both cases. The generative negative log-likelihood  $P(\text{obs}|\text{cluster})$  was used to measure the quality of model fit, with lower negative log-likelihood values (higher generative probabilities) indicating a better model fit. Fig. 2 shows the results. The median negative log-likelihood of the semi-Markov version is significantly better than that of the Markov version ( $p < 0.05$ , Wilcoxon-Mann-Whitney test). Comparison of the semi-Markov version of CDMC against a Hidden Markov Model (HMM) version (results not shown) also resulted in superior performance of the semi-Markov version ( $p < 0.05$ ).

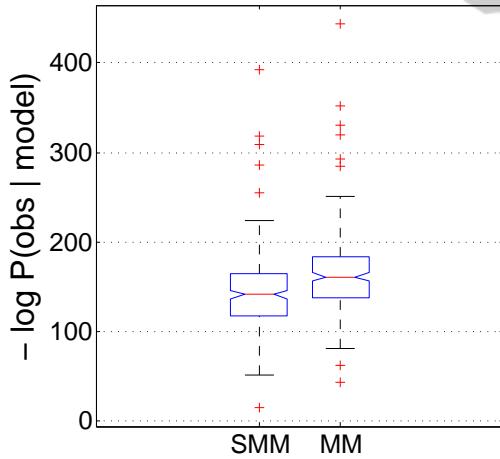


Figure 2: Negative generative log-likelihoods of CDMC clusters for semi-Markov (left) and standard Markov (right) dynamical models. Non-overlapping notches indicate significant difference in medians ( $p < 0.05$ ). Semi-Markov models provide significantly better log-likelihood.

Fig. 3 illustrates the CDMC clustering results for the semi-Markov dynamical models. The coordinates of each instance are the estimated parameter values of the Weibull distribution for that instance's wake duration distribution. Cluster centroids are significantly separated along both parameter axes ( $p < 0.05$ ).

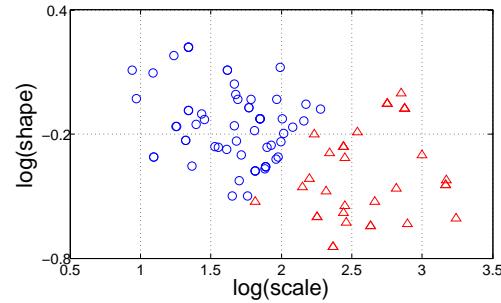


Figure 3: Visualization of CDMC clustering of human sleep data with Weibull semi-Markov dynamical model. Coordinates of each instance are parameter values of wake stage Weibull model fit individually to the instance.

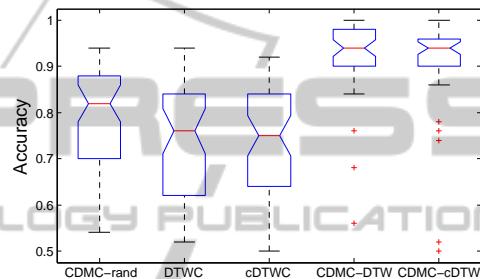


Figure 4: Accuracies for randomly initialized CDMC, DTW-only clustering (DTWC), constrained DTW-only clustering (cDTWC), DTW-initialized CDMC, and constrained DTW-initialized CDMC. Non-overlapping notches indicate significant difference in medians ( $p < 0.05$ ). DTW-initialized and cDTW-initialized CDMC yield significantly better accuracies than the other clustering techniques.

## 4.2 Initialization Technique

Pseudo-random initialization for CDMC was compared with DTW-distance clustering and Itakura constrained DTW (cDTW) clustering initializations for CDMC (section 2.3.2), as well as with the DTW and cDTW clustering results directly. Fig. 4 shows the resulting accuracy values. Pseudorandomly initialized CDMC (median accuracy 0.82) is not significantly more accurate than the two DTW-only clustering techniques without CDMC (medians 0.76 and 0.75, for full and constrained DTW, respectively). However, the DTW-initialized CDMC (median accuracy 0.94) significantly outperforms all other techniques. Furthermore, DTW initialization provides significantly faster convergence than pseudorandom initialization in the case of the Rand index as the similarity metric (left two boxes in Fig. 5.) Statistical significance is assessed at the  $p < 0.05$  level, using a Wilcoxon rank sum test (Mann-Whitney test).

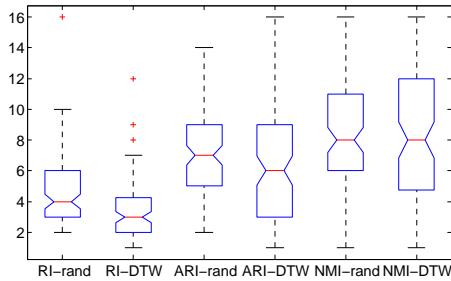


Figure 5: Iterations to convergence for different stopping criteria. Random and DTW-initialized CDMC. Semi-Markov mixture data. Stopping metrics: RI (two left), ARI (two center), NMI (two right). Non-overlapping notches indicate significant difference in medians ( $p < 0.05$ ). RI stopping is significantly faster than others. DTW initialization significantly speeds up convergence for RI stopping.

### 4.3 Stopping Criterion

The criterion for stopping the iterative process in CDMC depends on the clustering similarity metric. Three similarity metrics were compared (see section 2.3.3): the standard Rand index (RI), the adjusted Rand index (ARI) intended to correct for chance clustering agreements, and Normalized Mutual Information (NMI). The resulting numbers of iterations required for convergence are shown in Fig. 5. RI is seen to lead to a significantly lower median number of iterations to convergence (4 and 3, for pseudorandom initialization and DTW initialization, respectively) as compared with ARI (7 and 6 iterations) and NMI (8 iterations for both random and DTW initialization).

The faster convergence observed for RI as compared with ARI is in itself not surprising, as the numerical threshold  $\text{minSim}$  used for stopping in Algorithm 1 is the same for the three similarity metrics, while ARI has lower values than RI due to the intended correction of clustering agreement due to chance. One would therefore also expect that the more stringent ARI criterion would lead to better differentiated clusters. However, as Fig. 6 shows, the resulting median accuracy values from testing over labeled synthetic semi-Markov mixture data are not significantly different at the level  $p < 0.05$  (Wilcoxon-Mann-Whitney test). The combination of faster convergence and comparable accuracy points to RI as the superior choice of stopping criterion for CDMC.

## 5 CONCLUSIONS AND FUTURE WORK

The collective dynamical modeling-clustering (CDMC) algorithmic framework (Alvarez and Ruiz,

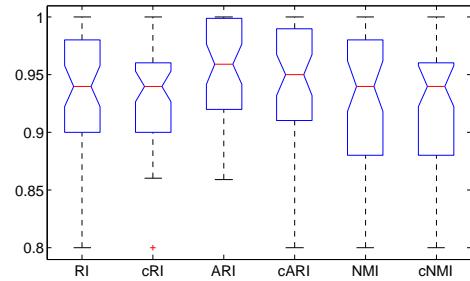


Figure 6: Accuracies for different stopping criteria of DTW-initialized and constrained DTW-initialized CDMC over semi-Markov mixture data. Stopping metrics: RI (two left), ARI (two center), NMI (two right). Non-overlapping notches indicate significant difference in medians ( $p < 0.05$ ). Median accuracy values do not differ significantly.

2013) is designed to more reliably identify rare dynamical events in sequence data by selectively aggregating instances based on dynamical similarity to increase sample size for modeling, simultaneously yielding a dynamics-based clustering.

The present paper uses semi-Markov chains as the CDMC dynamical models, as they better capture the dynamics of human sleep in comparison with the more widely used Markov models. Our experimental results over data from human sleep studies confirm the validity of this statement.

The use of distance-based dynamic time warping clustering for CDMC initialization is found in this paper to lead to significantly more accurate CDMC clustering results in experiments with labeled synthetic data than pseudorandom initialization does, as well as significantly faster CDMC convergence.

The adjusted Rand index was tested as the clustering similarity metric that defines the CDMC stopping criterion, in order to correct for clustering agreements due to chance. However, this was shown not to lead to significantly more accurate clusterings, while significantly increasing the number of iterations required for convergence as compared with the standard Rand index. Similar statements hold for the Normalized Mutual Information metric as compared with the standard Rand index. Therefore, the standard Rand index is the best choice of similarity metric for CDMC among these candidates.

One direction for future work is the use of partially observable semi-Markov models as the dynamical models.

## REFERENCES

- Alvarez, S. A. and Ruiz, C. (2013). Collective probabilistic dynamical modeling of sleep stage transitions. In *Proc. Sixth International Conference on Bio-inspired*

- Systems and Signal Processing (BIOSIGNALS 2013)*, pages 209–214, Barcelona, Spain. SciTePress.
- Berndt, D. J. and Clifford, J. (1994). Using dynamic time warping to find patterns in time series. In *KDD Workshop*, volume 10 (16), pages 359–370. Seattle, WA.
- Bianchi, M. T., Cash, S. S., Mietus, J., Peng, C.-K., and Thomas, R. (2010). Obstructive sleep apnea alters sleep stage transition dynamics. *PLOS ONE*, 5(6):e11356.
- Chervin, R. D., Fetterolf, J. L., Ruzicka, D. L., Thelen, B. J., and Burns, J. W. (2009). Sleep stage dynamics differ between children with and without obstructive sleep apnea. *Sleep*, 32(10):13251332.
- Chu-Shore, J., Westover, M. B., and Bianchi, M. T. (2010). Power law versus exponential state transition dynamics: application to sleep-wake architecture. *PLOS ONE*, 5(12):e14204.
- Hernandez, T. L., Ballard, R. D., Weil, K. M., Shepard, T. Y., Scherzinger, A. L., Stamm, E. R., Sharp, T. A., and Eckel, R. H. (2009). Effects of maintained weight loss on sleep dynamics and neck morphology in severely obese adults. *Obesity*, 17(1):84–91.
- Hubert, L. and Arabie, P. (1985). Comparing partitions. *Journal of Classification*, 2:193–218. 10.1007/BF01908075.
- Iber, C., Ancoli-Israel, S., Chesson, A., and Quan, S. (2007). *The AASM Manual for the Scoring of Sleep and Associated Events: Rules, Terminology and Technical Specifications*. Westchester: American Academy of Sleep Medicine.
- Itakura (1975). Minimum prediction residual principle applied to speech recognition. *IEEE Transactions on Acoustics, Speech, and Signal Processing*, 23(1):52–72.
- Kim, J., Lee, J. S., Robinson, P., and Jeong, D. U. (2009). Markov analysis of sleep dynamics. *Physical Review Letters*, 102(17):178104.
- Kishi, A., Struzik, Z. R., Natelson, B. H., Togo, F., and Yamamoto, Y. (2008). Dynamics of sleep stage transitions in healthy humans and patients with chronic fatigue syndrome. *American Journal of Physiology-Regulatory, Integrative and Comparative Physiology*, 294(6):R1980–R1987.
- Oates, T., Firoiu, L., and Cohen, P. R. (1999). Clustering time series with hidden Markov models and dynamic time warping. In *Proceedings of the IJCAI-99 Workshop on Neural, Symbolic and Reinforcement Learning Methods for Sequence Learning*, pages 17–21.
- Phillips, A. J., Robinson, P. A., Kedziora, D. J., and Abey-suriya, R. G. (2010). Mammalian sleep dynamics: how diverse features arise from a common physiological framework. *PLOS Computational Biology*, 6(6).
- Rand, W. M. (1971). Objective criteria for the evaluation of clustering methods. *Journal of the American Statistical Association*, 66(336):846–850.
- Rechtschaffen, A. and Kales, A. (1968). *A manual of standardized terminology, techniques and scoring system for sleep stages of human subjects*. US Department of Health, Education, and Welfare Public Health Service - NIH/NIND.
- Sakoe, H. and Chiba, S. (1978). Dynamic programming algorithm optimization for spoken word recognition. *IEEE Transactions on Acoustics, Speech and Signal Processing*, 26(1):43–49.
- Salvador, S. and Chan, P. (2007). Toward accurate dynamic time warping in linear time and space. *Intelligent Data Analysis*, 11(5):561–580.
- Vinh, N. X., Epps, J., and Bailey, J. (2010). Information theoretic measures for clusterings comparison: Variants, properties, normalization and correction for chance. *The Journal of Machine Learning Research*, 9999:2837–2854.
- Wang, C., Alvarez, S. A., Ruiz, C., and Moonis, M. (2013). Computational modeling of sleep stage dynamics using Weibull semi-Markov chains. In *Proc. Sixth International Conference on Health Informatics (HEALTHINF 2013)*, pages 122–130, Barcelona, Spain. SciTePress.
- Yu, S.-Z. (2010). Hidden semi-Markov models. *Artificial Intelligence*, 174(2):215 – 243. Special Review Issue.
- Zung, W. W., Naylor, T. H., Gianturco, D. T., and Wilson, W. P. (1965). Computer simulation of sleep EEG patterns with a Markov chain model. *Recent advances in biological psychiatry*, 8:335–355.