

Categorical Timeline Allocation and Alignment for Diagnostic Head Movement Tracking Feature Analysis

Mitsunori Ogihara
Dept. Computer Science
University of Miami
ogihara@cs.miami.edu

Zakia Hammal
Robotics Institute
Carnegie Mellon University
zhammal@andrew.cmu.edu

Katherine B. Martin
Tobii Pro
katherine.martin@
tobii.com

Jeffrey F. Cohn
Dept. Psychology
University of Pittsburgh
jeffcjohn@pitt.edu

Justine Cassell
Human-Computer Interaction Institute
Carnegie Mellon University
justine@cs.cmu.edu

Gang Ren
Ctr. Comp. Science
University of Miami
gxr467@miami.edu

Daniel S. Messinger
Dept. Psychology
University of Miami
dmessinger@miami.edu

Abstract

Atypical head movement pattern characterization is a potentially important cue for identifying children with autism spectrum disorder. In this paper, we implemented a computational framework for extracting the temporal patterns of head movement and utilizing the imbalance of temporal pattern distribution between diagnostic categories (e.g., children with or without autism spectrum disorder) as potential diagnostic cues. The timeline analysis results show a large number of temporal patterns with significant imbalances between diagnostic categories. The temporal patterns show strong classification power on discriminative and predictive analysis metrics. The long time-span temporal patterns (e.g., patterns spanning 15-30 sec.) exhibit stronger discriminative capabilities compared with the temporal patterns with relatively shorter time spans. Temporal patterns with high coverage ratios (existing in a large portion of the video durations) also show high discriminative capacity.

1. Introduction

Clinicians and researchers studying children with autism spectrum disorder (ASD) have long noticed the atypical head postures and movement patterns of children with ASD [1-3]. Previous findings are based on manual analysis where diagnostic cues are elicited from the domain experts based on their in-field experience [4-11]. Many of the related studies [4-8] are based on interactions with the subjects (e.g., children with and without ASD) in an interactive environment and then manually coding the videos captured. The interpretation stages in such workflows are usually focused on a few specific gestures or motion features chosen by domain experts. The chosen gestures, features, and the experimental environments are diverse in these studies, thus it is difficult to combine the techniques employed. Furthermore, many promising results obtained through such manual processes [6-8] are

difficult to scale up for larger studies [10, 11] due to the variation of the analytic devices, the experiment settings, and the interpretation procedures. The difficulty in integration and scaling up the studies necessitates the development of computational tools that assist the researchers at various stages of analyses and interpretations.

This paper reports a computational implementation of head movement tracking and analysis framework based on temporal pattern analysis and interpretation. Our proposed framework standardizes and integrates the head movement features reported in [1-11] while focuses on the temporal variation descriptors extracted from the dynamics of these head movement feature sequences. Fig.1 provides an illustration of the proposed system architecture. The approach consists of four steps. The first step extracts the head movement time series from the video. The second step performs temporal feature integration, which transforms the movement time series into token sequences for pattern analysis. The third step applies sequential pattern analysis tools to discover all frequent sequential patterns appearing in the head movement time series in each diagnostic category. The fourth step looks for sequential patterns with discriminative power among diagnostic categories. The patterns with high discriminative power are those that appear disproportionately in the two subject categories. By exploring a large number of discriminative sequential patterns, we are able to quantify their predictive connections to the diagnostic categories as potential cues for autism phenotypes identification, analysis, and interpretations.

The dataset, the participants, and the data collection protocol are covered in Sec. 2. Sec. 3 presents the feature extraction algorithms. Sec. 4 presents the temporal feature integration and organization steps. Sec. 5 covers the sequential pattern analysis and categorical timeline allocation algorithms. Sec. 6 presents the discriminative analysis metrics. Sec. 7 presents the analysis results and Sec. 8 provides a brief summary and discussions.

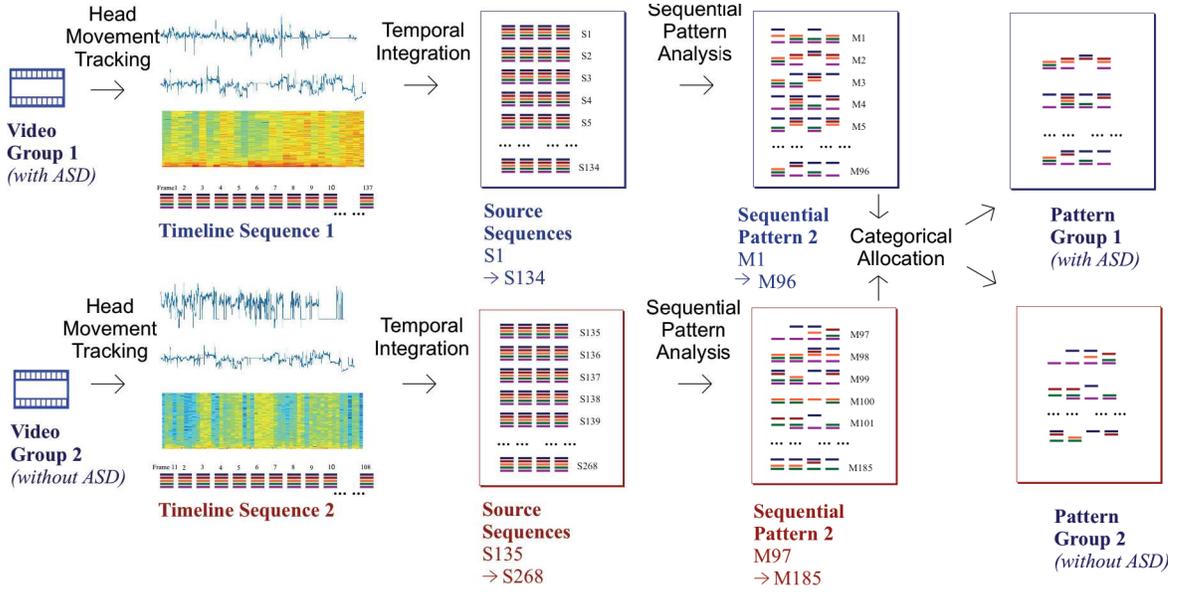


Fig. 1. System Architectural Overview. The head movement features are extracted from the videos using a head movement tracking software. Temporal feature integration and sequential pattern analysis algorithms extract the frequently recurring sequential patterns from the head movement feature sequences. Then the sequential patterns are mapped to the head movement feature sequences and grouped according to their subject categories.

2. Data Collection Protocol

Participants were children of three groups. The first group is children with ASD. The second group is children without ASD. They were typically developing children with no reported family history of ASD or ASD diagnosis. The third group is children without ASD who have one or more siblings with an ASD diagnosis [12]. Participants are between 2.5 and 6.5 years old, where mean = 4.72 years, standard deviation = 1.14 years, and range = 4.25 years, with ASD (n = 21), without ASD (n=21), without ASD but sibling(s) has ASD (n=14). The ADS diagnosis or the absence of ASD was determined by a licensed psychologist based on DSM-IV-TR criteria [13] and the Autism Diagnostic Observation Schedule (ADOS) [14]. The University's Internal Review Board approved recruitment procedures and experiment protocols.

Children were asked to watch six short videos on a computer monitor, while a camera positioned on top of the monitor recorded their face and upper body at 29.971 frame/s. The responses from these six videos form six video data blocks for each child.

3. Head Movement Feature Extraction and Modeling

Our implementation utilizes an automatic and person-independent face tracking system [15,16]. This face tracking system provides the rotation angle displacement of the head position at each video frame. The rotation angles include pitch (head nods), yaw (head turns), and roll (lateral head inclination). Our implementation also provides alternative head position

representations including the quaternions and the directional cosines. The quaternion-based head movement representation is calculated from the rotation angles as in [17]. The head movement representation based on quaternions form a four-dimensional vector space $[q_r(n), q_i(n), q_j(n), q_k(n)]$, where n is the index of the video frame, and $r, i, j,$ and k denote the four quaternion elements. These three movement representations are equivalent and can be transformed to each other as in [17-19]. We also calculate the first and the second temporal derivatives of head movement time series in each direction.

The velocity vector $\mathbf{V}(n) = [V_X(n) V_Y(n) V_Z(n)]$ is calculated from the derivatives of the quaternion q :

$$\begin{bmatrix} V_X(n) \\ V_Y(n) \\ V_Z(n) \\ 0 \end{bmatrix} = \frac{q(n+1) - q(n-1)}{\tau} \otimes \vec{q} \quad (1)$$

where \otimes denotes matrix multiplication. \vec{q} is the inverse of q . The scalar velocity is:

$$V_s(n) = \sqrt{V_X^2(n) + V_Y^2(n) + V_Z^2(n)} \quad (2)$$

The accelerations are calculated as:

$$A_X(n) = \frac{V_X(n+1) - V_X(n-1)}{2\tau} \quad (3)$$

The scalar acceleration is:

$$A_S(n) = \sqrt{A_X^2(n) + A_Y^2(n) + A_Z^2(n)} \quad (4)$$

The kinematic head movement feature sequences are transformed to the dynamic muscle actuation feature sequences depicting neck muscle activities using a head-neck simulation model [20-22]. This model provides a non-linear mapping relationship between the movement feature sequences and the corresponding muscle activation

signal. The muscle activation signal is then mapped to the muscle force using the Ornstein-Uhlenbeck process following [23]. This model assumes that the result of muscle activation is damped in time when mapped to the rotation angle kinematics. In other words, the velocity or the derivatives of rotation angles tends to return to zero unless there exist continuous muscle activations to sustain such movements. The parameters of the discretized motion model are fitted to the dataset using the autocorrelation method as described in [23,24]. This head-neck simulation model yields three head-neck muscle activation feature sequences corresponding to the three directions of rotation angles.

4. Temporal Feature Integration

Temporal feature integration provides additional feature descriptors that summarize the temporal variation patterns in the head movement feature sequences. In our analysis framework, temporal feature integration is performed both in the time domain and in the frequency domain.

4.1. Time Domain Integration

The temporal feature integration algorithm divides the time series data into short segments for extracting temporal integration descriptors. For each video data block, we choose the first 1200 data points (40 seconds in video duration). The length of the time segments for feature extraction is configured to four different lengths corresponding to a hierarchical time scale arrangement. The local analysis scale of the highest time resolution preserves the full resolution of the head movement feature sequences. The length of the time step is the video frames duration (0.033 sec.). The segment lengths of the other two local analysis scales divide the complete time series at each block to 50 segments and 20 segments. The global analysis scale divides the complete video duration into five equal-length segments.

For each segment of a time series, we calculate the mean value and the variance of the data points within this segment. Several temporal summarization parameters for longer time durations are calculated from the segment means and variances, which is defined as the followings:

- Average value of segment means

$$V_p = \frac{1}{P} \sum_{i=1}^P M(i) \quad (5)$$

where P is the number of short segments. $M(i)$ is the mean value of the data points in segment i .

- Average value of segment variances

$$V_m = \frac{1}{P} \sum_{i=1}^P V(i) \quad (6)$$

where $V(i)$ is the variance of the data points in segment i .

- Standard deviation of segment means

$$M_d = Std[M(1), \dots, M(P)] \quad (7)$$

where Std calculates the standard deviation of the following variables.

- Standard deviation of segment variances

$$M_d = Std[V(1), \dots, V(P)] \quad (8)$$

- Accumulative difference of segment means

$$M_a = \frac{1}{P-1} \sum_{i=1}^{P-1} |M(i+1) - M(i)| \quad (9)$$

where $|\cdot|$ denotes absolute value.

- Accumulative difference of segment variances

$$V_a = \frac{1}{P-1} \sum_{i=1}^{P-1} |V(i+1) - V(i)| \quad (10)$$

4.2. Frequency Domain Integration

For each short segment, we calculate the spectral centroid and span as the descriptors for the time-frequency energy distributions. Suppose that the discrete Fourier transform magnitude of the signal in short segment i is $S_i(f)$, where f is the frequency index, the spectral centroid of this time segment is calculated as the weight center of the spectral magnitude of the constituent frequency components:

$$C(i) = \frac{\sum_{f=1}^L f S_i(f)}{\sum_{f=1}^L S_i(f)} \quad (11)$$

where L is the total number of frequency bins. The spectral span $W(i)$ is calculated as the frequency point where the gross signal magnitude from the lower frequencies than this frequency point exceeds 70% of the gross signal magnitude in all frequencies. The temporal integration parameters is calculated as:

- Average value of the spectral centroids

$$C_m = \frac{1}{P} \sum_{i=1}^P C(i) \quad (12)$$

where i indices the short segments, and P is the total number of short segments.

- Average value of the spectral spans

$$W_m = \frac{1}{P} \sum_{i=1}^P W(i) \quad (13)$$

- Standard deviation of the spectral centroids

$$C_d = Std[C(1), \dots, C(P)] \quad (14)$$

where Std denotes the standard deviation of the following variables.

- Standard deviation of the spectral spans

$$W_d = Std[W(1), \dots, W(P)] \quad (15)$$

- Accumulative differences of the spectral centroids

$$C_a = \frac{1}{P-1} \sum_{i=1}^{P-1} |C(i+1) - C(i)| \quad (16)$$

where $|\cdot|$ denotes absolute value.

- Accumulative differences of the spectral spans

$$W_a = \frac{1}{P-1} \sum_{i=1}^{P-1} |W(i+1) - W(i)| \quad (17)$$

4.3. Organization of Feature Dimensions

Each feature dimension is a combination of the movement directions and the temporal integration options in Fig. 2. This combination results in 702 feature dimensions organized into four time resolution parts. Part 1 includes all global time scale feature dimensions that use five segments to depict the temporal variation of the full signal duration. For each movement direction, the temporal

integration options include the mean within segments of the five equal partitioned segments of the time series, the same segment means for the first derivative of the time series, the same segment means for the second derivative of the time series, the spectral centroid for the five segments, and the spectral width for the five segments. This adds up to 18 movement directions \times 5 temporal integration options = 90 feature dimensions. The series numbers of the feature dimensions are arranged as the first 5 feature dimensions correspond to movement direction 1, and then dimensions 6 to 10 correspond to the movement direction 2, etc.

Part 2 includes all local time scale feature dimensions with 20 temporal integration segments. For each movement direction, the temporal integration option first includes the mean within segments, the standard deviation within segments, the differences between successive segment means, the difference between successive segment standard deviations. Then these temporal integration options repeat for the first and the second derivatives of the time series. Finally the temporal integration options include the spectral centroid and the spectral width of the segments.

Part 3 includes all local time scale feature dimensions with 50 temporal integration segments. The feature dimension arrangement is identical to Part 2.

Part 4 includes the local time scale feature dimensions with the full time resolution (the original time series).

5. Sequential Pattern Analysis and Categorical Allocation

5.1. Feature Sequence Tokenization

The feature quantization step adapts the continuous feature values to discrete tokens for the subsequent sequential pattern analysis algorithm. Because the means and the variances of the feature values are already included as the temporal integration descriptors (as in Sec. 4), the feature quantization step in our implementation emphasizes the dynamic patterns (i.e., the change of feature values between successive time series steps) of the feature sequences instead of its static values. The feature quantization step calculates the differences of the feature values in successive time steps and quantizes the difference values into “shape” variables of “0” (decrease), “1” (hold), or “2” (increase). The difference values smaller than 10% of the standard deviation of all difference values of its corresponding feature dimension are quantized as token “1” (hold). A difference value is quantized as token “0” (decrease) or “2” (increase) if its absolute value exceeds 10% of the standard deviation of all difference values of its corresponding feature dimension and the variation is towards the decrease or increase direction respectively.

The template sampling algorithm chops the token sequences into shorter fragments. The template sampling

process in our implementation employs successive templates of length L_{temp} and the adjacent templates overlap $L_{temp} - 1$ steps. This is to keep the temporal context of L_{temp} steps for any token location without missing gaps between templates. This template sampling process chops a quantized feature sequence of length L into $L - L_{temp} + 1$ source sequences. The term “source sequence” is defined as a sequence fragment obtained from the template sampling process. At the global time scale level, the source sequences model the dynamics of the complete feature sequence. At the local time scale levels, source sequences model the local time dynamics within a short time duration centered at different time locations of the feature sequences.

| Feature Descriptors | | M1 | M2 | ... | M18 | |
|---------------------------------|-----|---------------------------------------|--------------|-------------|---------|-------------|
| Part 1 (5 Segments) | | 1-5 | 6-10 | 11-15 | 16-20 | |
| D1 | T1 | Mean Within Segment (D1, D2, D3) | 1,2,3 | 6,7,8 | ... | 86,87,88 |
| D2 | T4 | Spectral Centroid | 4 | 9 | ... | 89 |
| D3 | T5 | Spectral Width | 5 | 10 | ... | 90 |
| Part 2 (20 Segments) | | 91-104 | 105-118 | 119-132 | 133-146 | |
| D1 | T1 | Mean Within Segment (D1, D2, D3) | 91,95,99 | 105,109,113 | ... | 329,333,337 |
| D2 | T2 | Standard Deviation Within Segment | 92,96,100 | 106,110,114 | ... | 330,334,338 |
| D3 | T3 | Difference of Means (Successive Seg.) | 93,97,101 | 107,111,115 | ... | 331,335,339 |
| | T4 | Diff. of Std. (Successive Seg.) | 94,98,102 | 108,112,116 | ... | 332,336,340 |
| | T13 | Spectral Centroid | 103 | 117 | ... | 341 |
| | T14 | Spectral Width | 104 | 118 | ... | 342 |
| Part 3 (50 Segments) | | 343-356 | 357-370 | 371-580 | 581-594 | |
| D1 | T1 | Mean Within Segment (D1, D2, D3) | 343, 347,351 | 357,361,365 | ... | 581,585,589 |
| D2 | T2 | Standard Deviation Within Segment | 344,348,352 | 358,362,366 | ... | 582,586,590 |
| D3 | T3 | Difference of Means (Successive Seg.) | 345,349,353 | 359,363,367 | ... | 583,587,591 |
| | T4 | Diff. of Std. (Successive Seg.) | 346,350,354 | 360,364,368 | ... | 584,588,592 |
| | T13 | Spectral Centroid | 355 | 369 | ... | 593 |
| | T14 | Spectral Width | 356 | 370 | ... | 594 |
| Part 4 (Full Resolution) | | 595-600 | 601-606 | 607-696 | 697-702 | |
| D1 | T1 | Original Sequence (D1, D2, D3) | 595,597,599 | 601,603,605 | ... | 697,699,701 |
| D2 | T2 | Diff. Between Successive Points | 596,598,600 | 602,604,606 | ... | 698,700,702 |

(a)

| | | | | | |
|-----|----------------------|-------------------|-----|-----------------------------------|-------|
| M1 | Pitch Displacement | 1-5 | M12 | Head-Neck Muscle Activation Pitch | 56-60 |
| M2 | Yaw Displacement | 6-10 | M13 | Head-Neck Muscle Activation Yaw | 61-65 |
| M3 | Roll Displacement | 11-15 | M14 | Head-Neck Muscle Activation Roll | 66-70 |
| M4 | Directional Cosine 1 | 16-20 | M15 | Quaternion 1 | 71-75 |
| M5 | Directional Cosine 2 | 21-25 | M16 | Quaternion 2 | 76-80 |
| M6 | Directional Cosine 3 | 26-30 | M17 | Quaternion 3 | 81-85 |
| M7 | Nose Location X | 31-35 | M18 | Quaternion 4 | 86-90 |
| M8 | Nose Location Y | 36-40 | | | |
| M9 | Nose Location Z | 41-45 | | | |
| M10 | Velocity | 46-50 | | | |
| M11 | Acceleration | 51-55 | | | |
| | D1 | Original Sequence | | | |
| | D2 | 1st Derivative | | | |
| | D3 | 2nd Derivative | | | |

(b)

Fig. 2. Illustration of the temporal integration options and head movement feature directions for organizing the feature dimensions. The feature dimension index number can be looked up from the index number ranges in (a). The features are organized at four levels of temporal resolutions. Each resolution level includes temporal features and spectral features. All temporal features were analyzed with respect to derivative options D1, D2, and D3. Then both temporal and spectral features were applied to 18 movement feature directions (M1 - 18). (b) Illustrates the tokens of the movement feature directions and the derivative sequences. The feature indices in (b) provide the complete indices for “Part 1” features in (a) as an illustrative example.

5.2. Sequential Pattern Discovery

Sequential patterns are frequently recurring sub-sequences embedded in the source sequences [25]. In

our implementation, the sequential pattern discovery algorithm [26,27] is employed. This algorithm allows the user to specify several constraints that control the time distribution of gaps in the discovered sequential patterns and thus provides additional flexibility compared to conventional sequence analysis tools [25]. The algorithm admits a sub-sequence as a sequential pattern if all the itemsets in the non-gap locations in the sequential pattern occur in a subset of the source sequences (termed supporting source sequences) and the number of the supporting source sequences is larger than a preset threshold. The non-gap itemsets in a supporting source sequence must include the sequential pattern but the gap locations can contain any token. Admitting gaps relaxes the admission criterion of sequential patterns, allowing more long-length sequential patterns to be observed. Otherwise the long patterns with gaps will be split into shorter sequential patterns at the gap points. This procedure especially beneficial our implementation because it allows us to admit and evaluate longer sub-sequences as sequential patterns, which usually have higher contextual significance.

5.3. Categorical Sequence Grouping

The source sequences are partitioned into categorical groups according to their diagnostic categories corresponding to the feature sequences that they are extracted from. The categorical sequence grouping process identifies to occurrence number of each sequential pattern at each diagnostic categories. An illustrative example is presented in Fig. 3. Each row of the pattern alignment matrix [28] corresponds to a sequential pattern and each column corresponds to a source sequence. The alignment between a sequential pattern and source sequence is indicated by “1”. The source sequences are divided into three groups of ASD (“a”), without ASD (“n”), and sibling(s) has ASD (“s”). The total occurrence number of a sequential pattern in a diagnostic category is calculated as the total number of alignment in the corresponding source sequence group. For example, Patter 1 occurs 4 times in group “a”, 2 times in group “n”, and 4 times in group “s”. The combined diagnostic categories explored in this paper include “a+s” and “n+s”. The occurrence number of a sequential pattern in the combined diagnostic categories is calculated as the sum of the occurrence number in the constituent diagnostic categories.

6. Quantifying Discriminative Analysis Performances

6.1. Pattern Distributional Imbalance Metrics

The distributional imbalance of a sequential pattern across two categories is the main metric for its diagnostic power, which is quantified as the distributional contrast ratio (DCR) for each sequential pattern:

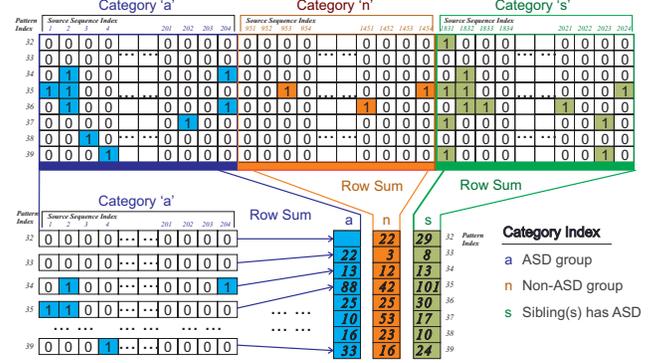


Fig. 3. Illustration of the pattern alignment matrix and the categorical sequence grouping process. The entries in the alignment matrix is “1” when the sequential pattern (indices of the rows) is included in the source sequence (indices of the columns), and “0” otherwise. The occurrence number of a sequential pattern in a diagnostic category is calculated as the sum of entries within the group of columns of this category.

$$DCR = \frac{P_1 \cdot P_2}{S_1 \cdot S_2} = \frac{P_1 S_2 - P_2 S_1}{P_1 S_2 + P_2 S_1} \quad (18)$$

where P_1 is the number of appearances of this sequential pattern in the source contour sequences from the Category 1. P_2 is the same number from the Category 2. S_1 is the number of source sequences in Category 1, and S_2 is the number in Category 2. The higher DCR value indicates a higher extent of imbalance in the appearance numbers of the discriminative sequential patterns between categories 1 and 2. By including S_1 and S_2 , this DCR definition also normalizes the effect of imbalanced source sequence numbers in different categories. The same sequential pattern tends to appear more in the category with more source sequences. So we use the coverage, instead of the appearance number, for calculating the DCR and related predictive analysis metrics. Therefore, all metrics reported to follow should be interpreted as the class-balanced case (each diagnostic category with the same number of source sequences).

The coverage of a sequential pattern within the source sequences of its category is calculated as:

$$CVG_1 = \frac{P_1}{S_1} \quad (19)$$

$$CVG_2 = \frac{P_2}{S_2} \quad (20)$$

A higher coverage value means the corresponding sequential pattern can perform diagnostic classification for more source sequences, thus more effective if the predictive measures are identical.

6.2. Predictive Analysis Metrics

Several predictive analysis parameters can be calculated from the DCR values by treating each sequential pattern as a classifier. For example, if a sequential pattern has large DCR value at the “a→n” direction, the appearance of this sequential pattern in a new (unseen) feature sequence is a strong evident that it comes from the “a” category. Suppose a sequential pattern appears P_1 times in the

source sequences in category 1 and P_2 times in category 2. Then we split the features (existence of the sequential pattern) and labels (diagnostic category) into two parts as non-overlapping training set and test set. The probability that there are more category 1 source sequence than category 2 source sequence in the training set (this sequential pattern indicating that the source sequences containing it is category 1) is:

$$P_{s1} = \frac{P_1}{P_1+P_2} \quad (21)$$

The estimated mean value of category 1 source sequence in the independent test set is:

$$M_{s1} = \frac{P_1}{P_1+P_2} \cdot N_{te} \quad (22)$$

where N_{te} is the number of source sequences in the test set.

Taking Category 1 as positive category, the true positive (TP) number is calculated as:

$$TP = P_1 \cdot M_{s1} = \frac{P_1^2}{(P_1+P_2)^2} N_{te} \quad (23)$$

The true negative (TN) number is

$$TN = \frac{P_2^2}{(P_1+P_2)^2} N_{te} \quad (24)$$

The false positive (FP) and false negative (FN) number are:

$$FP = FN = \frac{P_1 P_2}{(P_1+P_2)^2} N_{te} \quad (25)$$

The accuracy value is

$$ACC = \frac{TP+TN}{TP+TN+FP+FN} = \frac{P_1^2+P_2^2}{(P_1+P_2)^2} = \frac{DCR^2+1}{2} \quad (26)$$

The precision value is:

$$PRE = \frac{TP}{TP+FP} = \frac{P_1}{P_1+P_2} = \frac{DCR+1}{2} \quad (27)$$

The recall value is:

$$REC = \frac{TP}{TP+FN} = \frac{P_1}{P_1+P_2} = \frac{DCR+1}{2} \quad (28)$$

The F1 score is:

$$F1 = \frac{2 \cdot REC \cdot PRE}{REC+PRE} = \frac{P_1}{P_1+P_2} = \frac{DCR+1}{2} \quad (29)$$

The Cohen's Kappa is:

$$\kappa = 2 \cdot ACC - 1 = DCR^2 \quad (30)$$

6.3. Discriminative Sequence Ranking and Selection

A subset of the sequential patterns are selected as the discriminative sequence and then utilized for calculating the discriminative and predictive metrics. Our implementation provides three options for selecting the discriminative sequence with the emphasis on DCR, coverage, and their combination.

The first pattern selection option sort the patterns according to their DCR values in descending order for each feature dimension and contrast pair. Then the first or the last ten patterns are selected as the discriminative patterns. The pattern selection process is illustrated in Fig. 4. Each rectangular box represents a sequential pattern. The DCR value is plotted inside the box and the index number of the pattern is plotted below the box. This example is based on an "a/n" contrast pair and all DCR

values are calculated in the "a→n" direction. The mean value of the DCR values is calculated as the sum of the absolute values from the top 10 patterns. When selecting the patterns for the "n/a" contrast pair, the ten patterns in the negative end of the sorting chain are selected. The mean value of the absolute values of the DCR values then just flips the negative sign to positive for the mean DCR values for this example. We note that the mean DCR values are calculated as the mean of the absolute values because the positive/negative sign of a DCR value only indicate the pattern distribution imbalance direction, while the absolute value indicting the discriminative power. Because DCR values in both the two opposite directions (e.g., "a→n" and "n→a") show the same values for discriminating the contrasting diagnostic categories (e.g., "a" and "s"), in our implementation, we also provide a combined bi-directional "a↔n" direction, as the combined "a→n" and "n→a" directions, by collecting the top patterns with the highest absolute DCR values from both ends of the DCR sequence.

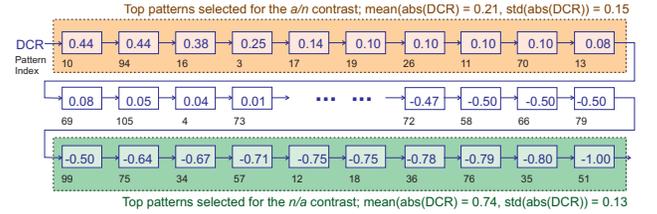


Fig. 4. The first pattern selection option for calculating the average DCR value for a pair of diagnostic categories. The DCR values in this example is calculated as the a→n contrast and sorted in descending order. The top ten patterns selected for the a/n contrast will be different from the top ten patterns selected for the n→a contrast when there are more than ten patterns in the sorted sequence. The n→a contrast direction will also flip the negative/positive values when calculating the statistics. The combined a↔n direction takes DCR values of the highest absolute values from both end of the sorted DCR sequence.

The second pattern selection option selects ten sequential patterns with top coverage values in the first diagnostic category of the contrast pair and then calculates the same metrics. This pattern selection option put more emphasis on the pattern coverage over the diagnostic discriminative capacity.

The third pattern selection option seeks a balance point between the DCR values and the coverage when selecting sequential patterns. Within each feature dimension and contrast pair, this option first selects 40 sequential patterns with the highest coverage, then from these 40 patterns selects ten sequential patterns with the highest DCR values.

7. Results and Discussion

Figures 5 through 11 present the summary graphs of categorical timeline allocation results across all feature dimensions and all contrast pairs. The organization of the feature dimensions is presented in Sec. 4.3. These figures present the distributional patterns of the average DRC values and the average coverage values of the sequential

patterns in each categorical contrast pairs and/or feature dimensions. A higher average DRC value indicates that the sequential patterns in the corresponding feature dimension and contrast pair have higher discriminative power. A higher average coverage value means that the corresponding sequential patterns can discriminate more subjects and is thus more useful. The contrast pairs include the diagnostic categories among “a”- children with ASD, “n”- children without ASD, “s”- siblings with ASD, and two combined groups of “a+s” and “n+s”. The category “s” means that the features are extracted from children who do not have ASD but their sibling(s) have ASD. For each contrast pair, the subjects on the left of the “→” sign are termed Category 1 and the subjects on the right are termed Category 2. The double side arrow “↔” indicating combined contrast directions.

Fig. 5 present the average DCR values for the ten sequential patterns with the highest DCR values for each feature dimension and each contrast pair for the first pattern selection option. The color of each entry corresponds to the values as shown in the colorbar on the right. The color towards deep blue is mapped to the low values, while dark red as high values. We also provided grayscale and other color mapping versions as electronic companion material for this paper. Using grayscale mapping, the dark color indicates low value and the light color indicates high value. The series number of feature dimensions can be looked up in Fig. 2. Each entry (“pixel”) in Fig. 5 corresponds to the DCR values of a feature dimension and a contrast pair. For example, the top-leftmost entry is from feature dimension 1 (pitch displacement, means within segments) and contrast pair ASD/non-ASD. The four areas marked under the figure correspond to the four temporal integration parts of feature dimensions as illustrated in Sec. 3 and Fig. 2. Area 1 (global time analysis scale) has significantly higher DCR values compared to the other three areas (local time analysis scale). Among Areas 2, 3 and 4, Area 2 (local time analysis scale with 20 segments) has slightly higher DCR values in most parts compared to Area 3 (local time analysis scale with 50 segments). Area 3’s DCR values are larger than the DCR values in Area 4 (local time analysis scale with full time resolution). Among the contrast pairs, “a→s”, “n→s”, “a↔s”, and “n↔s” have the largest DCR values.

For the predictive analysis metrics, Fig. 6 presents the corresponding F1 scores. The F1 scores are a linear mapping from the DCR values with different dynamic ranges (more details on the predictive analysis metrics in Sec. 6.2). For each single DCR values, the dynamic range of [0, 1] in DCR is linearly mapped to [0.5, 1] in the F1 scores. The precision and the recall metrics are identical with the F1 score. Fig. 7 presents the average Kappa values. The DCR/Kappa transformation is nonlinear but monolithic. The accuracy metric follows a similar trend as the Kappa distribution.

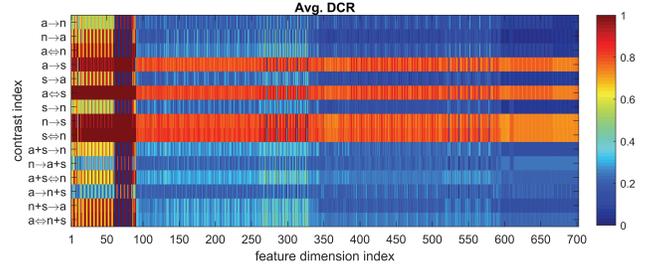


Fig. 5. Average DCR values for the ten sequential patters with the highest DCR values for each feature dimension and each contrast pair. The colors in the grid indicate the average DCR value as shown in the color bar. High DCR values indicate higher discriminative power. The four areas correspond to the time analysis scales as illustrated in the text.

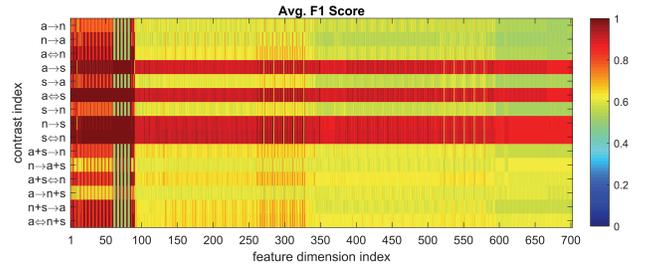


Fig. 6. Average F1 scores for the ten sequential patters with the highest F1 scores for each feature dimension and each contrast pair. The distribution of F1 scores are more towards “1”.

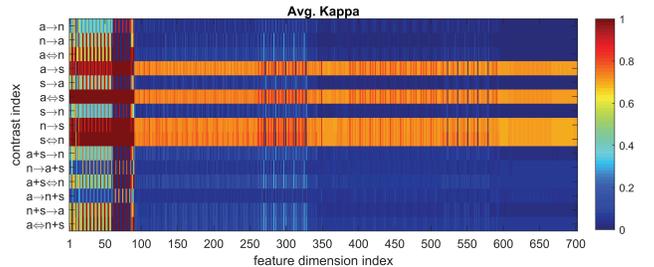


Fig. 7. Average Kappa values for the ten sequential patters with the highest F1 scores for each feature dimension and each contrast pair. The distribution of F1 scores are more towards “1”.

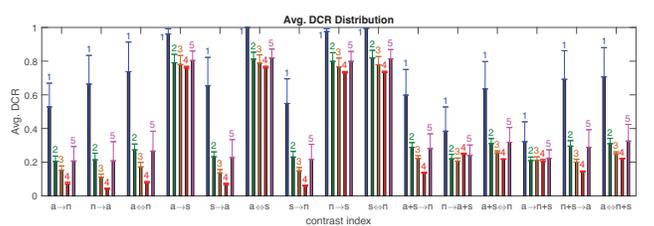


Fig. 8. Area summary of the DCR distribution pattern. For each contrast pair, we calculate the mean (lengths of the bars) and the standard deviation (twice the lengths of the errorbars) within the four areas in Fig. 5 (marked as “1”, “2”, “3”, and “4”) and all areas (marked as “5”).

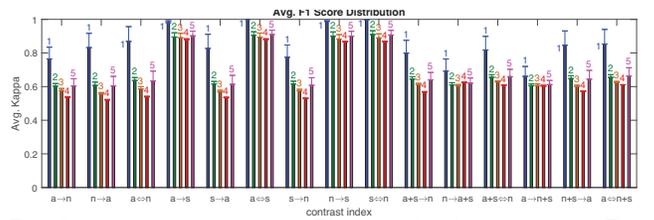


Fig. 9. Area summary of the F1 score distribution pattern. The distribution of F1 scores are more towards “1”.

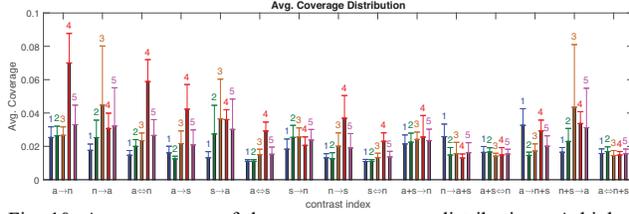


Fig. 10. Area summary of the average coverage distributions. A higher coverage value means that the corresponding pattern can discriminate more subjects and thus more useful.

Fig. 8 presents the area summary of the DCR distribution pattern in Fig. 5. For each contrast pair, we calculate the mean (lengths of bars) and the standard deviation (twice the lengths of the errorbars) within each area (marked as “1”, “2”, “3”, “4” and “5” near the errorbars; “5” as all areas). For all contrast pairs, the average DCR values are highest in Area 1, then gradually decrease from Areas 2, 3, and 4. We can also observe that the contrast pairs with the highest DCR values are “a→s” and “n→s”. Contrast pairs “a+s→n” and “n+s→a” have high DCR values in the rest part of the categorical pairs. The corresponding bi-directional combined pairs “a↔s”, “n↔s”, “a+s↔n”, and “n+s↔a” also shows high DCR values, as higher DCR values for either direction (e.g., “a→s” and/or “s→a”) lead to the high DCR values in the combined direction (“a↔s”). Fig. 9 presents the area summary of the F1 score distribution pattern, which shows the identical patterns as in Fig. 8.

Fig. 10 presents the average coverage values for the ten sequential patterns with top DCR values for each feature dimension and each contrast pair. Area 1 (global time scale) has the lowest coverage value in most feature dimensions. The contrast pairs of “a→n”, “n→a”, “a+s→n”, “n+s→a”, “a↔n”, “a+s↔n”, and “n+s↔a” have higher average coverage values in all areas compared to the rest contrast pairs.

The results from the second pattern selection option (more details in Sec. 6.3) are presented in Fig. 11 and 12. This option results in lower DCR values compared to the previous setting of selecting ten patterns with the highest DCR values. Still Area 1 has the highest DCR values and the categorical pairs of “a→s”, “n→s”, “a+s→n”, “n+s→a”, “a↔s”, “n↔s”, “a+s↔n”, and “n+s↔a” have higher DCR values. The average coverage values are much higher because only top ten patterns with the highest coverage values in each feature dimension and each contrast pair are considered.

The results from the third pattern selection option are presented in Figures 13 and 14. This option yields moderate DCR and coverage values. The DCR part (Fig. 13) shows similar pattern as in the previous two options. The coverage (Fig. 14) show high coverage values for the contrast pairs of “a→n”, “n→a”, “a→s”, “n→s”, “a+s→n”, and “n+s→a”, as well as the corresponding combined pairs of “a↔n”, “a↔s”, “n↔s”, “a+s↔n”, and “n+s↔a”.

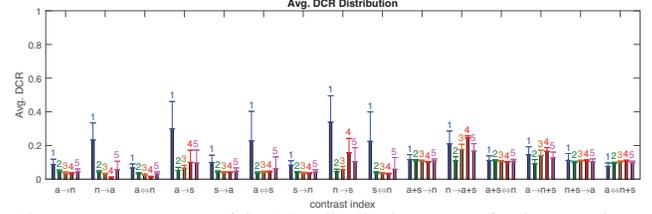


Fig. 11. Area summary of the DCR distribution pattern for the second pattern selection option, which selects ten sequential patterns with top coverage values.

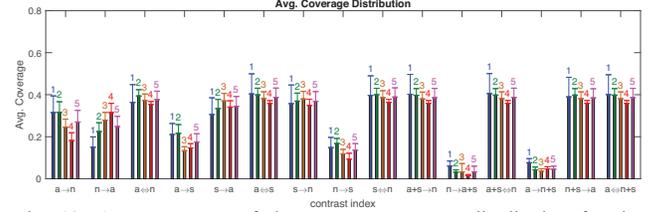


Fig. 12. Area summary of the average coverage distribution for the second pattern selection option.

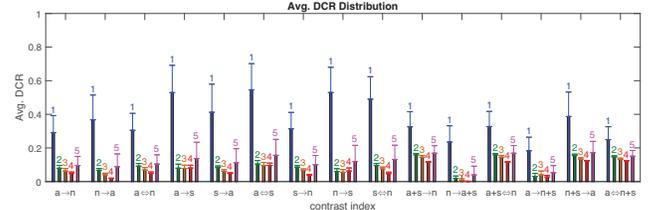


Fig. 13. Area summary of the DCR distribution pattern for the third pattern selection option, which selects ten sequential patterns with top DCR values within 40 sequential patterns of higher coverage.

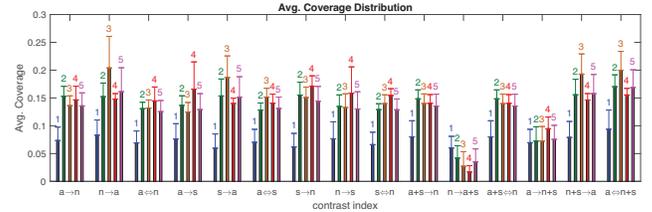


Fig. 14. Area summary of the average coverage distribution for the third pattern selection option.

8. Summary and Conclusion

We implemented a computational framework for extracting the head movement feature sequences from multiple motion directions, performing temporal pattern analysis from these feature sequences, and analyzing dominant temporal patterns from different diagnostic categories. The proposed framework is applied on a video dataset of head movement in response of watching various video stimuli for autism spectrum disorder (ASD) research. The categorical pattern allocation process have identified the temporal pattern with significant imbalanced distributions among diagnostic categories including children with ASD, children without ASD, and children without ASD but their sibling(s) have ASD. A large number of temporal patterns show significantly imbalanced distributions from different diagnostic categories at most feature dimensions and temporal resolution levels.

References

- [1] Goldman, S., et al. Motor stereotypies in children with autism and other developmental disorders. *Developmental Medicine & Child Neurology*, **51**(1): p. 30-38, 2009.
- [2] Fournier, et al. Motor coordination in autism spectrum disorders: a synthesis and meta-analysis. *Journal of Autism and Developmental Disorders*, **40**(10): p. 1227-40, 2010.
- [3] Freeman, B.J., et al. The Behavior Observation Scale for Autism: Initial methodology, data analysis, and preliminary findings on 89 children. *Journal of the American Academy of Child Psychiatry*, **17**(4): p. 576-588, 1978.
- [4] Zwaigenbaum, L., et al. Behavioral manifestations of autism in the first year of life. *International Journal of Developmental Neuroscience*, **23**(2-3): p. 143-152, 2005.
- [5] Kim, S.H. and C. Lord. Restricted and repetitive behaviors in toddlers and preschoolers with autism spectrum disorders based on the autism diagnostic observation schedule (ADOS). *Autism Research*, **3**(4): p. 162-173, 2010.
- [6] Ozonoff, S., et al. A prospective study of the emergence of early behavioral signs of autism. *Journal of the American Academy of Child & Adolescent Psychiatry*, **49**(3): p. 256-266, 2010.
- [7] Mottron, et al. Lateral glances toward moving stimuli among young children with autism: Early regulation of locally oriented perception? *Development and Psychopathology*, **19**(1): p. 23-36, 2007.
- [8] Hellendoorn, A., et al. The relationship between atypical visual processing and social skills in young children with autism. *Research in Developmental Disabilities*, **35**(2): p. 423-428, 2014.
- [9] Piek, J.P. and M.J. Dyck. Sensory-motor deficits in children with developmental coordination disorder, attention deficit hyperactivity disorder and autistic disorder. *Human movement science*, **23**(3): p. 475-488, 2004.
- [10] Ozonoff, S., et al. Gross motor development, movement abnormalities, and early identification of autism. *Journal of Autism and Developmental Disorders*, **38**(4): p. 644-656, 2008.
- [11] Chang, C.-H., et al. Visual tasks and postural sway in children with and without autism spectrum disorders. *Research in Developmental Disabilities*, **31**(6): p. 1536-1542, 2010.
- [12] Messinger, D., Young, G. S., Ozonoff, S., Dobkins, K., Carter, A., Zwaigenbaum, L., Landa, R. J., Charman, T., Stone, W. L., Constantino, J. N., Hutman, T., Carver, L. J., Bryson, S., Iverson, J. M., Strauss, M. S., Rogers, S. J., & Sigman, M. Beyond Autism: A Baby Sibling Research Consortium Study of High-Risk Children at Three Years of Age. *Journal of the American Academy of Child and Adolescent Psychiatry*, **52**(3), 300-308. PMC3625370, 2013.
- [13] Lord C, et al. The autism diagnostic observation schedule—generic: a standard measure of social and communication deficits associated with the spectrum of autism. *J Autism Dev Disord*, **30**(3):205–23, 2000.
- [14] Lord C, Rutter M, Le Couteur A. Autism diagnostic interview-revised: a revised version of a diagnostic interview for caregivers of individuals with possible pervasive developmental disorders, *J Autism Dev Disord*, **24**(5): 659–85, 1994.
- [15] <http://zface.org>
- [16] L. A. Jeni, J. F. Cohn and T. Kanade. "Dense 3D face alignment from 2D videos in real-time," *2015 11th IEEE International Conference and Workshops on Automatic Face and Gesture Recognition (FG)*, Ljubljana, pp. 1-8, 2015.
- [17] Jazar, Reza N. *Theory of applied robotics: kinematics, dynamics, and control*. New York: Springer, 2010.
- [18] Marghitu, Dan B. *Mechanisms and robots analysis with MATLAB*. London, UK: Springer, 2009.
- [19] Corke, Peter. *Robotics, Vision and Control: fundamental algorithms in matlab*. New York: Springer, 2017.
- [20] Zangemeister, W.H., Lehman, S. & Stark, L. Simulation of head movement trajectories: model and fit to main sequence, *Biol. Cybern.* **41**: 19, 1981.
- [21] Zangemeister, W.H., Jones, A., Stark, L.: Dynamics of head movement trajectories: main sequence relationship. *Exp. Neurol*, **71**: 76–91, 1981.
- [22] Abrahams, V. The physiology of neck muscles; their role in head movement and maintenance of posture. *Can. J. Phys*, **55**, 332–338, 1975.
- [23] J. J. Heuring and D. W. Murray. "Modeling and copying human head movements," in *IEEE Transactions on Robotics and Automation*, vol. 15, no. 6, pp, 1095-1108, Dec 1999.
- [24] B. Xiao, P. Georgiou, B. Baucom and S. S. Narayanan. "Head Motion Modeling for Human Behavior Analysis in Dyadic Interaction," in *IEEE Transactions on Multimedia*, vol. 17, no. 7, pp, 1107-1119, July 2015.
- [25] M. J. Zaki. "SPADE: An Efficient Algorithm for Mining Frequent Sequences", *Machine Learning Journal*, Vol. 42(1), pp, 31-60, 2001.
- [26] Y. Hirate and H. Yamana. "Generalized sequential pattern mining with item intervals," *Journal of Computers*, vol. 1(3), pp, 51–60. 2006.
- [27] P. Fournier-Viger. SPMF: A Java Open-Source Data Mining Library. [Online]. Available: <http://www.philippe-fournier-viger.com/spmf/>. Accessed on: Mar. 18, 2019.
- [28] T. Rakthanmanon, B. Campana, A. Mueen, G. Batista, M. Westover, Q. Zhu, J. Zakaria and E. Keogh: "Searching and mining trillions of time series subsequences under dynamic time warping," in *Proc, KDD*, Beijing, China, pp. 262-270, 2012.