Sequential Pattern Mining of Multimodal Data Streams in Dyadic Interactions

Damian Fricker Hui Zhang Chen Yu

Indiana University, Bloomington, IN 47408, U.S.A
{dfricker, huizhang, chenyu}@indiana.edu

***Abstract*—In this paper we propose a sequential pattern mining method to analyze multimodal data streams using a quantitative temporal approach. While the existing algorithms can only find sequential orders of temporal events, this paper presents a new temporal data mining method focusing on extracting exact timings and durations of sequential patterns extracted from multiple temporal event streams. We present our method with its application to the detection and extraction of human sequential behavioral patterns over multiple multimodal data streams in human-robot interactions. Experimental results confirmed the feasibility and quality of our proposed pattern mining algorithm, and suggested a quantitative data-driven way to ground social interactions in a manner that has never been achieved before.**

***Index Terms*—Cognitive Science, Human Robot Interaction, Sequential Pattern Mining.**

I. INTRODUCTION

The real-time behavioral dynamics of collaborating social partners in human-human interactions are composed of sequences of coordinated bodily cues -- rapid shifts of eye gaze, head turns, and hand gestures -- that happen on a time scale of fractions of a second. It is those coordinated behaviors that make the whole interaction appear to be so smooth and effortless. However, a closer look of those behaviors [1, 2] has revealed that such interactions are indeed highly complex on two grounds: 1) such interactions are multimodal, as we interact with each other through multiple communication channels (e.g., eye-gaze, speech, and pointing); and 2) such behaviors are adaptive, but not pre-defined, as people actively perceive information from their social partners and dynamically adjust their own actions to form real-time perception-action loops within one agent and across two agents. For example, in natural circumstances, the eye, head, and hand within an agent are in continual motion in the context of ongoing behavior [14] wherein eye, head, and hand all need to act with respect to a common coordinate system and remain synchronized in time across multiple actions. For another example, in face-to-face interactions, the speaker always looks at the listener's face at the end of a spoken utterance with an expectation that the listener would produce a back-channel cue (e.g. nodding or mutual gaze) to confirm the reception of the speech [15]. In addition, this confirmation action needs to happen at the right moment in time; otherwise the same action generated with the wrong timing, such as with an unexpected delay, may disrupt communication.

Analyzing real-time behavioral data in natural communications, how those behaviors are coordinated at critical timing moments, and the sequential patterns they exhibit, is critical to advancing our understanding of human-human communication as well as human-robot interactions. New sensing and computational technologies – eye-tracking systems, and advanced techniques for storing and automatically annotating massive data streams – promise new insights of a micro-level understanding of multimodal behaviors that aligns with contemporary explanations of cognitive and neural processes. In light of this, the goal of our research is to focus on data-mining and studying the sequential patterns exhibited in multimodal perception-action loops during real-time interactions within an agent and between two agents, i.e., how what one looks at or says relates to what the same person is going to look at or say next, as well as how the behavior of one agent influences what the other agent will look at and do next.

One particular challenge here is to discover reliable sequential patterns and the precise timing of free-flow human behaviors when they interact with another human or a robot in naturalistic contexts. Distinct from a robot executing a pre-defined action, human behaviors are much more adaptive and flexible. At a micro-behavioral level, it is rarely the case that two instances of the same human action are exactly executed in the same way in both time and space even though they may share underlying patterns (e.g. the trajectories of two instances of the same reaching action). Toward our research goal, we develop a sequential pattern mining algorithm to analyze multiple event-based data streams with the focus of discovering quantitative timings between sequential events from multiple temporal streams. We first describe the algorithm followed by applying it with multimodal data collected from a real-time human robot interaction platform that we built. We demonstrate the potential utilities of this method by providing several concrete examples of how this method is used to discover various kinds of micro-level behavioral patterns in human-robot interactions.

978-1-61284-990-4/11/$26.00 ©2011 IEEE

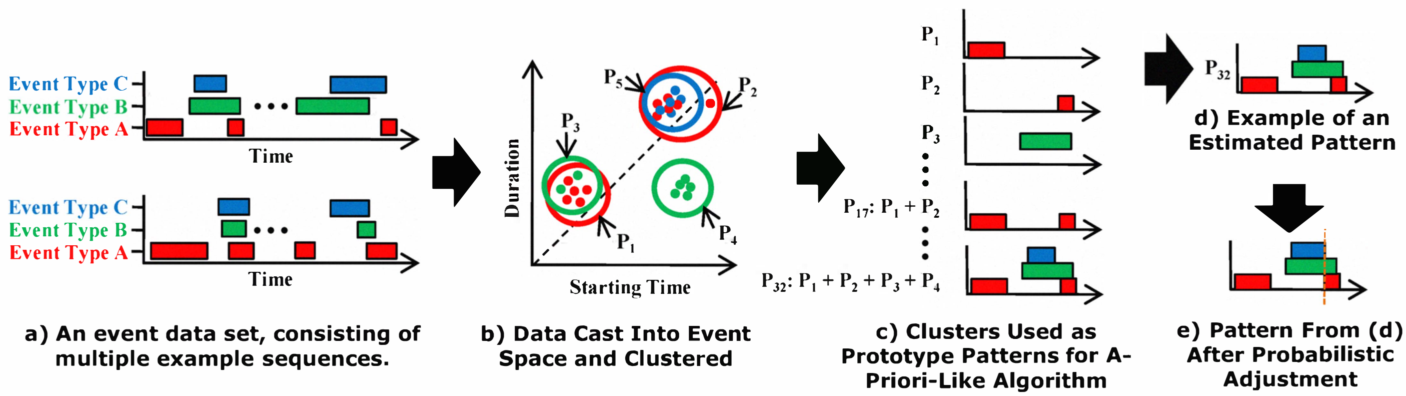

Figure-1: A process flow diagram of the algorithm. (a) An example subset of the event data set, displaying two example sequences, each with three event types. This is what is used as input to the algorithm. (b) This first step of the algorithm casts each example of the event data set into points in 2D space (event start time x event duration) and clusters the points. (c) Each cluster is used as an intial canadite pattern by an iterative Apriori-like algorithm. (d) Example of one estimated pattern detected from the Apriori-like algorithm. (e) A final adjustment to refine and adjust the temporal relations by removing some potential noise from the data.. Whether or not this is used is based on top-down knowledge of the data set, i.e., whether or not you would expect events to be completely synchronys.

II. METHOD

With huge amounts of multimodal data collected from various studies of human-human and human-robot studies, there have already been previous research efforts focused on issues such as gesture spotting and recognition [3], temporal pattern analysis [4], and predictions of human behavior patterns [5]. Most existing work employed temporal pattern mining methods developed for point-based events, which dates back to Agrawal and Srikant in 1995 [6]. Although those data-driven approaches provided useful means to understand multimodal interactions in human-human and human-robot interactions, an intrinsic limitation was that multimodal data streams in both human-human and human-robot interactions are indeed interval-based events instead of point-based events. Human behavioral patterns are composed of sequences of coordinated social cues on time scales of fractions of seconds; therefore point-based events are not able to represent events with this time scale and such complexity.

A. Related Work

Interval-based event mining was first proposed by Kam and Fu as an algorithm to discover temporal patterns for interval-based events (see [7]). Similar to point-based event mining algorithms, interval-based event mining algorithms use an Apriori-like algorithm to find longer frequent patterns from examples. As Allen discussed the difference between time points and time intervals (see [8]), interval-based event mining is essentially different from point-based event mining in that interval-based events are time-stamped with continuous starting and ending timestamps instead of discrete time points by which point-based events are represented.

However most interval-based event mining algorithms developed since then only focus on finding temporal orders between events [10-13] (e.g., event A is happening before/after event B). Quantitative timing information, such as at what moment a temporal event happens with what timing, has been a missing component in various data-mining and data-analysis scenarios. To our best knowledge, [9] is the only work to identify continuous interval-based events using a quantitative temporal mining algorithm. The approach exploited Apriori algorithm to generate candidate sequences as the first step, then a MCMC sampling method and an expectation-maximization (EM) algorithm were applied to select candidate patterns. This approach however has several issues. First, its time complexity is significantly high due to the sampling method and the EM algorithm. Second, the hypercube their approach constructs is based on the occurrences of events, which significantly limits its expressional flexibility. Lastly, this approach does not consider the relationship between events. In the next section, we propose our own approach to address these issues.

B. Continuous Interval-based Event Mining

In this section, we propose a new algorithm, event space miner (ESM), used to data-mine interval-based event patterns from example sets. An event data set consists of multiple example sequences, all of the same total duration, which can contain any number of event types, e.g., an event data set of natural disasters could contain multiple examples, each a year long, with timing information, example sequences, of volcanic eruptions, earthquakes, and tsunamis. Given an event data set, our ESM consists of 3 steps: 1) First, the algorithm converts the data into an event-space (see Figure-1(c)) and clustering is performed in the event space based on event types; 2) Second, these clusters become the first set of candidate patterns for an Apriori-like procedure to compute representative sequential prototypes (see Figure-1(d)); 3) Those prototype patterns are refined by considering their temporal relationships (see Figure-1(e)). Each step in the algorithm is described in details in the following sections.

1) Clustering: Given example sequences, ESM converts the examples into points in 2D vectors embedded in the constructed 2D event space (e.g., see Figure-1(c)), for EM algorithm to find clusters in each event type. We take the center points of the clusters as seed patterns for an Apriori-like algorithm. Each seed pattern is a 2D vector. Different from classic EM algorithm where the number of Gaussians is predefined, we only define the maximum number of Gaussians to be executed before we find the best fitting number using a model selection method -- normalized entropy criterion (NEC)[1].

[1]We used the package from www.mixmod.org

2) Apriori-like Algorithm: Typically Apriori-like algorithms first scan the database to find all the frequent single items, which are the initial input used as seed patterns, e.g. $P_1 \ldots P_i$. Then they generate 1-item longer candidate sequences which are tested if they are more frequent than a given threshold. If they are frequent enough, they will be used as seed patterns for the next iteration. The search for candidate sequence is performed recursively until there is no new frequent pattern or no candidate can be generated. This final output is a set of patterns, ($\{P_1{:}t_1\text{-}t_2\}$, $\{P_1{:}t_1\text{-}t_2;\ P_2{:}t_3\text{-}t_4\}$), where $t_\#$ is a timestamp.

Our customized Apriori-like algorithm first uses the mean vectors of the clusters as the candidates (e.g., see Figure-1(d)). Our algorithm then examines the occurrence frequency of each candidate against a threshold and re-generates new candidates in a new round by excluding those candidates that appeared less frequently. We note that our algorithm calculates the frequency of a pattern in a fuzzy way based on how much the pattern overlaps the examples. Let p be a pattern and x be an example event with starting time x.s and ending time x.e, the frequency is calculated by

$$f(p) = \frac{1}{|S|}\sum_{x \in S} h(p,x),$$

where $|S|$ is the size of example set and

$$h(p,x) = \max\left\{\frac{\min(p.e, x.e) - \max(p.s, x.s)}{p.e - p.s}, 0\right\}.$$

In this way, if an example overlaps with the prototype pattern only 10% of the time, then that example still adds 0.1 to the overall frequency counting.

3) Probabilistic Adjustment: Given an event data set, ESM converts the data into event-space and finds frequent patterns. However, the temporal relationships between the events have not been considered in pattern discovery. Given the examples as in Figure-1(b), which were generated by adding Gaussian noise to a prototype, the estimated pattern is shown in Figure-1(d) wherein three temporal events on the right seem to happen at slightly different time points due to noises in the examples.

We improved the results using the probability distributions of the events. Each event in the estimated pattern has its own distribution from examples. We assume that the distributions are Gaussian. We then consider the starting time and the ending time of an event separately as in point-based event mining algorithms. The procedure is given as follows:

1. Calculate the mean and variance for each point (a starting or ending time of each event).
2. Sort the distances between each pair of the means.
3. Check the probability that the closest pair happens actually at the same time. If the probability is higher than a given threshold,
   a. Adjust all the connected points.
   b. Update the means and variances of all the connected points.
   c. Register the pair into the connected list.
4. Go to 3 unless the next closest pair's distance is bigger than a threshold.

More specifically, each point has the mean and the variance from the examples, which are enough to estimate its Gaussian distribution. Let G_1 and G_2 be the two Gaussian distributions for two points from two different events, respectively. They are described by

$$G_1 \sim N(\mu_1, \sigma_1^2),$$
$$G_2 \sim N(\mu_2, \sigma_2^2).$$

Then, the probability that the distance between the two means is zero is given by

$$p(\text{distance} = 0) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{-\frac{1}{2\sigma^2}\mu^2\right\},$$

where $\sigma^2 = \sigma_1^2 + \sigma_2^2$ and $\mu = \mu_1 + \mu_2$, assuming that the two distributions are independent (or uncorrelated). If they are correlated, $\sigma^2 = \sigma_1^2 + \sigma_2^2 - \sigma_{12}^2 - \sigma_{21}^2$, where σ_{12}^2 and σ_{21}^2 are the cross-covariance.

Given two sets of examples with G1 and G2 distributions and with N1 and N2 numbers of examples, if the probability of zero distance is higher than a given threshold, it means they are close enough to be considered as the same point. The threshold is set to 0.2 in the experiments here. If they are close enough, they can be combined and make a new distribution as follows. Once they are combined, they move together as one point with new frequency score, mean and standard deviation:

$$N_{com} = N_1 + N_2,$$
$$\mu_{com} = \frac{N_1}{N_{com}}\mu_1 + \frac{N_2}{N_{com}}\mu_2,$$
$$\sigma_{com}^2 = \frac{1}{N_{com}}\left(N_1(\sigma_1^2 + \mu_1^2) + N_2(\sigma_2^2 + \mu_2^2)\right) - \mu_{com}^2.$$

In this way, those two events are merged as a new composed sequential event in Figure-1(e). This is an optional step that requires some top-down knowledge of the data set, i.e., should events be completely synchronous, is one type of event causally related to the prior event? One possible direction of future work is to estimate the parameters in this step based on the timing distribution of the overall event data set.

III. Experiment and Data

In this section, we introduce our human-robot interaction experimental paradigm. Figure-2 shows the overview of our multimodal real-time human-robot interaction. In a word-learning task we designed, a human teacher was asked to teach a robot the names of a set of novel objects in a shared environment. Participants were allowed to freely point, hold and manually manipulate those objects to attract the robot's attention and then name those objects using the artificial names we provided (e.g. "tema" and "dodi"). The robot is equipped with two cameras located in the forehead with capabilities of providing 640x480 resolution images at 30 fps. In particular, the robot was able to perform gaze-contingent actions by following a human participant's gaze in real time (e.g., mutual gaze, attentional object following, nodding/shaking etc).Using this platform, we have designed and conducted several experimental conditions in which the robot may sometimes follow the human's attention or sometimes ignore the human's attention by looking at a spatial location to where the human was not attending. In addition, the robot may sometimes look more toward the human's face to initialize mutual gaze or sometimes look less at the human's face. Our research goal was to record and analyze responsive behaviors from participants when they interacted with the same robot but with different gaze-contingent behaviors.

Multimodal data streams were recorded from both human participants and the robot in those experimental conditions, including human participants' eye gaze data, speech acts, first-person view video, and as well as the robot's body movements and head/gaze direction. We have developed various video, speech and motion data processing tools to automatically annotate raw multimodal data. As a result, we derived a set of temporal event streams from raw data which is used for the following sequential pattern discovery.

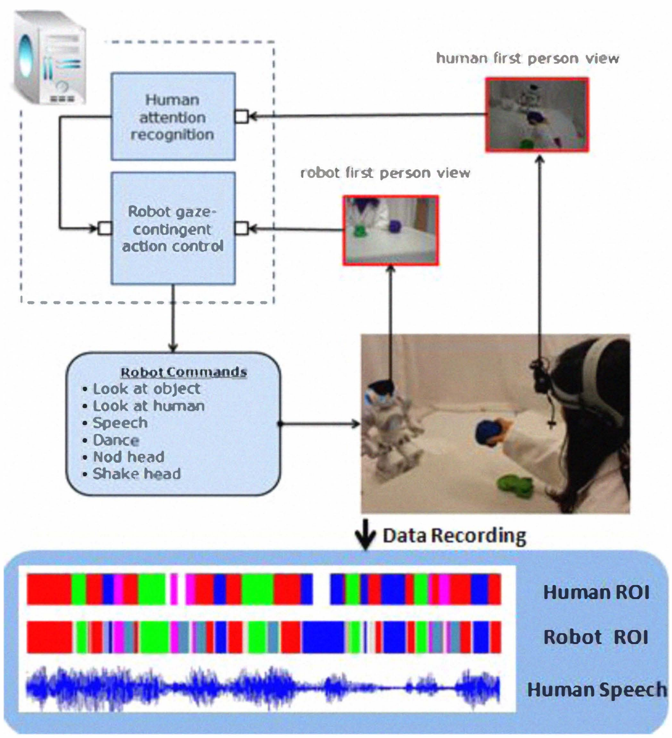

Figure-2: An overview of system structure. A participant and the robot sat across the table and interacted with each other in a shared environment. The human teacher attempted to teach the robot a set of (artificial) objects to the robot, and the robot generated gaze-contingent responsive behaviors based on the real-time detection of the human teacher's attention. In addition, multiple data streams were recorded to analyze human behaviors in this joint task.

IV. Sequential Patterns in Multimodal Data Streams

Our primary goal in this section is to test and demonstrate that ESM can discover various kinds of meaningful patterns from multimodal human-robot interaction data. Toward this goal, we focus on four main types of sequential patterns that capture multimodal perception-action dynamics both within human participants and between participants and the robot: one agent/unimodal, one agent/multimodal, two-agent/ unimodal, and two-agent/multimodal. As we mentioned earlier and demonstrate with multiple examples below, a key feature that distinguishes ESM with other existing algorithms is that we are able to extract exact timings and durations of sequential patterns from multiple data streams.

The set of temporal event streams comes from 15 subjects, each subject's data consisting of a total of 8 minutes across 4 conditions. In order to specify the questions we want to answer, we use top-down knowledge to chunk the data streams at key moments, e.g., to answer what are the quantitative timings to patterns of gaze around naming events, we would chunk the data streams into 5 second sequences based on each naming event and use those sequences as the event sequences of the event data set.

1) Unimodal Patterns from a single agent: One agent unimodal patterns can reveal sequential patterns within a single stream and how those patterns may evolve over time. In our human-robot interaction task, there are several unimodal streams, e.g., human speech, human gaze, and robot gaze. Here, we select human speech as an example. The speech data was transcribed and then coded as one of three types of speech acts: naming an object, describing an object, and attention getting. Naming events were sentences like "This is a wawa," in which an object name was mentioned in a spoken utterance; Describing events include those spoken sentences like "The wawa is blue and looks kind of like a 'y'." in which participants described physical properties of the object to the robot; Attention getting speech events were those that participants used speech to attract the robot's attention, such as "hi, robot, look over here."

Figure-3 shows the two most reliable patterns of speech events detected by ESM: 1) a 1.4 second naming sentence followed by a 3-second describing event with a 400 ms silence in between; 2) a shorter naming followed by a describing event which is then followed by an attention getting act. These sequential patterns reveal two key different strategies being used by participants in teaching the robot object names. In fact, we found that the first pattern most likely appeared in the beginning of the interaction to provide the key information in this teaching task, while the second pattern was used more often later as a re-iteration of the information with attention getting speech acts as a confirmation or a query of the information being received at the robot's end. In addition, a shorter duration of naming events (600 ms) in the second pattern indicates that participants were likely to just use an object name alone in a naming utterance without other spoken words in the later part of teaching.

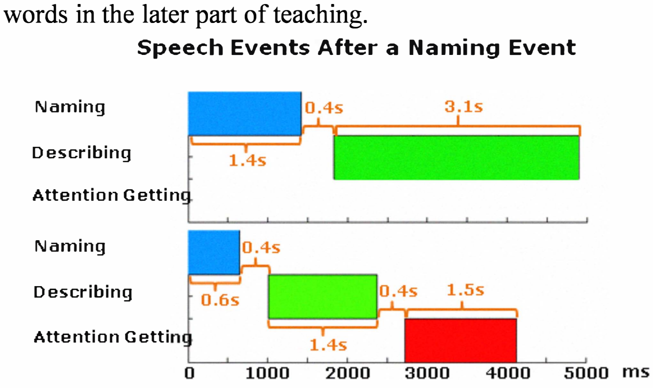

Figure-3: The two most reliable sequential patterns found in human's speech data stream. (a) A sequential pattern showing a 1.4s naming utterance followed by a 3s describing utterance. (b) A sequential pattern showing a 0.6s naming utterance followed by a 1.4s describing utterance and a 1.4s attention getting utterance. Note how all of the pauses between utterances across both patterns last about 0.4 seconds.

2) Multimodal patterns from a single agent: Multimodal patterns within an agent can reveal the interplay and synchronization of multimodal action, i.e., how gaze fixations on a target object and the naming of that object co-occur over time. To demonstrate the usefulness of ESM to extract these kinds of sequential patterns, we select three event streams from participants -- human naming events, human object looking and human face looking -- and fed them into ESM. Figure-4shows the most reliable sequential pattern detected by ESM data miner: Prior to a naming event (around 500 ms), participants gazed at the to-be-named object first, and then started a naming event while looking at the target object. Around 500ms after the onset of the naming event, participants quickly checked the robot's face for 700 ms and then switched their gaze back to the named object for 1.5 seconds.

This sequential pattern from human-robot interactions is in line with two previous findings of human behaviors. First, psycholinguistic studies in speech production showed that participants were likely to gaze at the object that will be referred to in the following spoken utterance before the onset of that spoken utterance to facilitate sentence planning [16]. Second, the most likely moment that a participant checked the robot's face was right after a naming utterance was started. This result is supported by [15], which found that during the onset of speech, humans tend to look away from the listener as they process the information required to utter the sentence. A brief look on the partner's face serves as a communicative signal to continue communication or to finish their social turn in a conversation. Thus, those confirmed results based on literature demonstrated that ESM is able to successfully discover sequential patterns that are embedded in the data.

3) Unimodal patterns across two-agents: We next take the human gaze stream and the robot gaze stream to analyze joint attention behaviors between two social partners. This analysis allows us to identify coupled and adaptive behaviors between two agents to not only discover those joint attention moments wherein they paid attention to the same region-of-interests but also who was leading who into a joint state. In particular, we focused on face gazing behaviors to ask how the human and the robot jointly attend to each other's face and what are the temporal differences between a human leading and robot leading mutual gaze.

Figure-5 shows the most reliable mutual gaze pattern detected from two gaze streams. Overall, the robot's face gazing duration was significantly longer than human's face

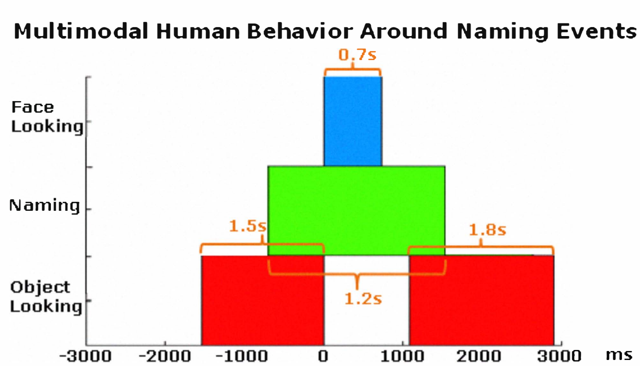

Figure-4: The most reliable sequential pattern detected from human's speech and gaze streams. Participants gazed at the to-be-named object before naming and then checked the robot's face 500 ms after the onset of the naming event followed by a 1.5-second look at the target object.

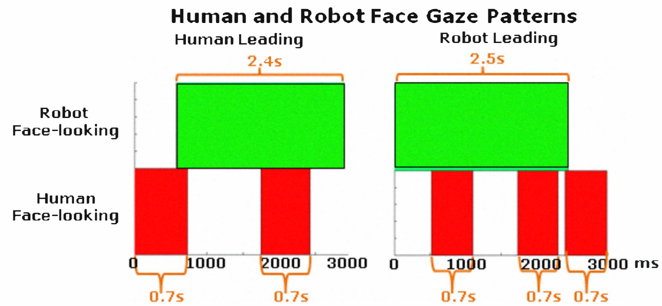

Figure-5: Two mutual gaze patterns discovered: (a) robot leading (b) human leading. Moreover, the results showed the robot's face gaze duration was significantly longer than human's face gaze events (a) and (b) humans produced more face looks.

gazing events and humans produced more face looks than the robot. Moreover, we found two different ways to lead to mutual gaze between humans and the robot: human leading (left) and robot leading (right).

In the human leading pattern (left), the results show that the robot was following the human's face gazing after 0.6 seconds, which was comparable to the amount of the time it took the human to follow a face gazing from the robot (that is, the robot leading pattern on the right). From the human perspective, a participant must first notice that the robot is looking at him, decide to look back, and then plan a saccade to switch his attention to the robot's face. This whole process seems to take around 600ms in human-robot interactions. In our platform, it was our original design to program the robot's gaze following behaviors to emulate human gaze behaviors. Therefore, the similar timing delay in the human following and the robot following cases verified the real-time aspects of our human-robot interaction platform. In addition, we found that a continued face looking from the robot (right) made participants look back the robot's face periodically; showing human's sensitivity to the robot behaviors and their adaptive behaviors to further facilitate interaction.

*4) Multi-modal patterns across multiple-agents: Multi-*agent multimodal patterns are the most impenetrable (and arguably most interesting) aspects of sensorimotor dynamics within multimodal interaction, since they encompass the dynamics both across modalities as well as across agents. Here, we select five temporal streams: the robot's face looking, the human's face looking, the robot's object gazing, the human's object gazing and an object moving event indicating whether the target object is moving or not. Figure-5 shows the most reliable pattern that ESM discovered from five event streams. At the beginning, the robot looked at the human's face while the human looked at a moving target object and then the robot switched its attention to the target object by following the human's attention. The robot's object following behavior triggered the human to check the robot's face which in turn led the robot to look to the human's face as a response action of the human's face looking. After that, the human has switched back to the target object and stays focused on it, even while the robot continued to look at the human's face -- both went back to their initial state after this sequence of responsive actions. We also found that right at the moment after the human switched to the robot's face (presumably to check the robot's attention), the human seemed to stop moving the target object. One plausible reason is that it might be easier for the human to check whether the robot was following at a non-moving target. Also, in this sequential pattern detected, we found two instances of the robot following of the human's gaze. The first can be seen as the robot followed the human gaze to the target object within about 1 second after the onset of the human gaze. The second instance of the robot following was the moment that the robot followed the gaze to the face about 0.5 seconds after the onset of the human gaze. More generally, this complex cascading sequence happening within a short window of time reflects the real-time dance of multimodal interactions, which we argue is the key to understand both human-human communication and human-robot interaction. Therefore, ESM is able to successfully discover rather complex patterns across multiple data streams, showing its potential utilities in analyzing complex micro-level behavioral data.

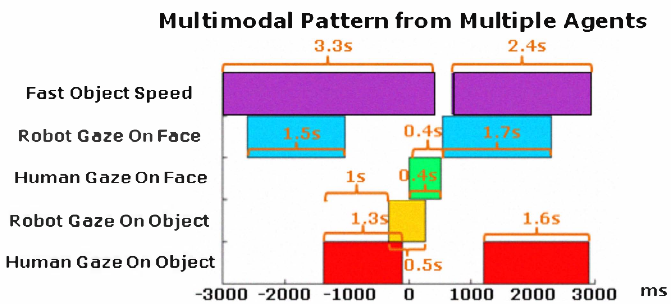

Figure-6: An example of sequential patterns from multiple events showing that both humans and the robot dynamically adapt their behaviors based on their real-time perception of their social partner's behaviors.

V. CONCLUSION

Discovering sequential patterns from multimodal data is an important topic in various research fields, such as human-human communication, human-agent and human-robot interactions, and human development and learning, as those fine-grained patterns can advance our understanding of human cognition and learning, and also provide quantitative evidence that we can directly incorporate in human-robot interactions. While the existing algorithms can only find sequential orders of temporal events, this paper presents a novel, not pre-defined, temporal data mining method focusing on extracting exact timings and durations of sequential patterns extracted from dense datasets of multiple temporal events. While top-down knowledge of the dataset is used to determine how to chunk the streams of data, the patterns themselves are attained in an unsupervised way. Using a multimodal human-robot interaction dataset, our experiment results showed that ESM data mining algorithm was able to detect and validate various kinds of reliable temporal patterns from multi-streaming, multi-modal data streams. More importantly, we demonstrated that the ESM algorithm can not only quantify patterns in sequential events, but also demonstrates the ability to find meaningful patterns across multiple modalities and across two interacting agents to capture different kinds of sensorimotor dynamical patterns in the interaction.

In light of this, this temporal data mining method has a potential to be useful in multiple research topics. Starting from the human-robot interaction framework and the sequential pattern mining algorithm, we plan to proceed to extend and continue our current research in various multimodal social scenarios, from human-robot interaction to child-parent interaction. We argue that understanding the coupling and sequential patterns from micro-level human behaviors can lead to a mechanistic account of fundamental aspects of human communication and interaction, and exact timings of multimodal behaviors within an agent and across two social partners can play a critical role in such smooth interaction.

ACKNOWLEDGMENT

We thank Hong-Wei Shen and Amanda Favata for help with running subjects, Computational Cognition and Learning Lab members for help in coding data, and Henry Choi and Hong-Wei Shen for programming assistance. Research was supported by NSF BCS 0924248 and AFOSR FA9550-09-1-0665.

REFERENCES

[1] Chen Yu, Matthias Scheutz, and Paul Schermerhorn, "Investigating multimodal real-time patterns of joint attention in an HRI word learning task," in *Proceedings of the 5th ACM/IEEE international conference on Human-robot interaction*, 2010.

[2] Hui Zhang, Damian Fricker, Thomas G. Smith, and Chen Yu, "Real-time adaptive behaviors in multimodal human-avatar interactions," in *International Conference on Multimodal Interfaces and the Workshop on Machine Learning for Multimodal Interaction*, 2010.

[3] Hee-Deok Yang, A-Yeon Park, Seong-Whan Lee, "Gesture Spotting and Recognition for Human–Robot Interaction," *IEEE Transactions on Robotics*, vol.23, no.2, pp.256-270, April 2007.

[4] Min Chen, Shu-Ching Chen, Mei-Ling Shyu, and K. Wickramaratna, "Semantic event detection via multimodal data mining, "*IEEE Signal Processing Magazine*, vol.23, no.2, pp.38-46, March 2006.

[5] N. Kubota and K. Nishida, "Prediction of Human Behavior Patterns based on Spiking Neurons for A Partner Robot," in *The 15th IEEE International Symposium on Robot and Human Interactive Communication*, 2006, pp.692-697.

[6] R. Agrawal and R. Srikant, "Mining sequential patterns," in *Proceedings of the 11th International Conference on Data Engineering*, 1995, pp. 3–14.

[7] P. Kam and A. W. Fu, "Discovering temporal patterns for interval-based events," in *Proceedings of the 2nd International Conference on Data Warehousing and Knowledge Discovery*, 2000.

[8] J. F. Allen, "Maintaining knowledge about temporal intervals," *Communications of the ACM*, 1983, vol. 26, no. 11.

[9] T. Guyet and R. Quiniou, "Mining temporal patterns with quantitative intervals," in *Proceedings of the 4th International Workshop on Mining Complex Data*, 2008.

[10] J. Han, J. Pei, B. Mortazavi-Asl, Q. Chen, U. Dayal, and M. Hsu, "Freespan: Frequent pattern projected sequential pattern mining," in *Proceedings of the 6th ACM/SIGKDD International Conference on Knowledge Discovery and Data Mining*, 2000, pp. 355–359.

[11] M. Mouhoub and J. Liu, "Managing uncertain temporal relations using probabilistic interval algebra," in *IEEE International Conference on Systems, Man, and Cybernetics*, 2008, pp. 3399-3404.

[12] P. Papapetrou, G. Kollios, S. Sclaroff, and D. Gunopulos, "Discovering frequent arrangements of temporal intervals," in *Proceedings of the 5th IEEE International Conference on Data Mining*, 2005, pp. 354-361.

[13] D. Patel, W. Hsu, and M. L. Lee, "Mining relationships among interval-based events for classification," in *Proceedings of the ACM SIGMOD International Conference*, 2008, Pp. 393-404.

[14] M. Hayhoe and D. Ballard, "Eye movements in natural behavior," *Trends in Cognitive Sciences*, 2005, vol. 9, no. 4, pp. 188-193

[15] M. Argyle, and M. Cook, *Gaze and Mutual Gaze*, London: Cambridge University Press, 1976.

[16] A. S. Meyer, A. Sleiderink, and W. J. M. Levelt, "Viewing and naming objects: Eye movements during noun phrase production, "*Cognition*, 1998, vol. 66, B25–B33