A Data-Driven Paradigm to Understand Multimodal Communication in Human-Human and Human-Robot Interaction

Chen Yu, Thomas G. Smith, Shohei Hidaka, Matthias Scheutz, Linda B. Smith

Psychological and Brain Scienecs and Cognitive Science Program,
1101 East 10th Street, Indiana University, Bloomington, IN, 47405
{chenyu}@indiana.edu

Abstract. Data-driven knowledge discovery becomes a new trend in various scientific fields. In light of this, the goal of the present paper is to introduce a novel framework to study one interesting topic in cognitive and behavioral studies -- multimodal communication between human-human and human-robot interaction. We present an overall solution from data capture, to data coding and validation, and to data analysis and visualization. In data collection, we have developed a multimodal sensing system to gather fine-grained video, audio and human body movement data. In data analysis, we propose a hybrid solution based on visual data mining and information-theoretic measures. We suggest that this data-driven paradigm will not only lead to breakthroughs in understanding multimodal communication but also more generally serve as a successful case study to demonstrate the promise of data-intensive discovery which can be applied in various research topics in cognitive and behavioral studies.

Keywords: Scientific Discovery, Cognitive and Behavioral Studies, Human-Human Interaction, Human-Robot Interaction, Information Visualization, Data mining

1 Introduction

With advances in computing and sensing technologies, the dominant methodology of science is changing over years. [1] (also see [2]) predicted that the first three paradigms in science – empirical, theoretical and computational simulation – have successfully carried us to where we are and will continue to make incremental progress, but meanwhile dramatic breakthroughs will be achieved by the next fourth paradigm of science – data-intensive science, which will help bring about a profound transformation of scientific research. In brief, a vast volume of scientific data captured by new instruments in various labs is likely to be substantially publically accessible for the purposes of continued and deeper data analysis. This analysis will result in the development of many new theories from such data mining efforts. Indeed, data-driven discovery has already happened in various research fields, such as earth sciences, medical sciences, biology and physics, to name a few. However, cognitive and behavioral studies still mostly rely on traditional experimental paradigms (reviewed below). The goal of the present paper is to introduce a contemporary framework to study one interesting topic in behavioral studies -- multimodal communication between human-human and human-robot interaction. We present an overall solution from data capture, to data coding and validation, and to data analysis and visualization. We suggest that this data-driven paradigm will not only lead to breakthroughs in understanding multimodal communication but also more generally serve as a successful case study to demonstrate the promise of this data-intensive approach which can be applied in many other research topics in cognitive and behavioral studies.

Everyday human collaborative behavior (from maintaining a conversation to jointly solving a physical problem) seems so effortless that we often notice it only when it goes awry. One common cognitive explanation of how we manage to (typically) work so well together is called “mind-reading” [3]. The idea is that we form models of and make inferences about the internal states of others; for example, along the lines of “He is looking at the object and so must want me to pick that up.” Accordingly, previous empirical methods on human-human communication are rather limited. For example, survey-based methods have been widely used to study human social interaction. This kind of measure relies on participants to recall and self-report their experiences and although these reports may be predictive and diagnostic, they need not be objectively correct and thus are at best an imperfect indication of what makes for “good” versus “not good” social interactions. Another popular approach is based on video coding in which human researchers code and interpret video data of human-human everyday interaction based on the prior notions about what is worth counting. But this rather subjective method may confirm what we already know but overlook important aspects of social interactions *that we do not yet know* – the ultimate goal of scientific discovery.

It is not at all clear that mind-reading theories about the states of others – and inferences from such internal representations – can explain the real-time smooth fluidity of such collaborative behaviors as everyday conversation or joint action. The real-time dynamics of the behaviors of collaborating social partners involve micro-level behaviors, such as rapid shifts of eye movements, head turns, and hand gestures, and how they co-organize across partners in an interaction. These behaviors seem to be composed of coordinated adjustments that happen on time scales of fractions of seconds and that are highly sensitive to the task context and to changing circumstances. Previous survey and video-coding approaches don’t have access to such fine-grained behavioral data, not to say interpret such micro-level behaviors. An understanding of micro-level real-time behaviors, however, is critical to building effective teams that can solve problems effectively, to building better social environments to facilitate human-human communication, to helping people that have various communication problems (e.g. autism), to building artificial agents (intelligent robots, etc.) that work seamlessly with people through human-like communication, and to building social training contexts for people to learn better (classroom interaction between teachers and students). Indeed, a fundamental problem in understanding both natural and artificial intelligent systems is the coordination of joint activity between social partners.

In both human-human communication and human-robot interaction, there is growing interest in more micro-analytic studies of just what happens – in real time – as individual agents interact. Within this new trend, an understanding of human collaboration requires a level of analysis that concentrates on sensory-motor behaviors as a complex dynamic system in which the behaviors of social partners continually adjust to and influence each other. Although there is a growing consensus for the need for such an approach, there has been little progress. The limits to progress include issues concerning how to measure fine-grained behaviors in real-time interactions, and how to analyze, quantify, and model the results.

This paper presents a contemporary framework to study real-time multimodal communication between autonomous agents (humans or robots). In the following sections, we will introduce a set of novel solutions under this data-driven paradigm, including how to collect high-resolution behavioral data from multimodal interaction (Section 2), how to code and mange the whole dataset (Section 3), and how to analyze and visualize the data to discover new patterns and principles in both human-human and human-robot interaction (section 4). With this set of data capture, data coding and data analysis techniques, our framework allows us to study critical questions that cannot be asked before using traditional methods, to discover new knowledge that is unlikely to be acquired using traditional paradigms, and to demonstrate the power of this data-driven approach to cognitive and behavioral fields.

2. Data Collection – A Multimodal Sensing System

The first component in a complete data-driven paradigm is data acquisition. Toward this end, we have developed a multimodal sensing environment in which we ask two agents (humans or robots) to interact with each other with pre-defined communication tasks. We have successfully used this experimental setup to collect high-resolution data from three interaction scenarios (each with its own specific research goals): adult-adult interaction to capture the fundamental principles in human-human communication, child-parent interaction to study social environments of developing children, human-robot interaction to discover behavioral patterns that the robot should emulate in order to perform human-like interaction. For instance, we asked parents to teach children a set of novel object names, and we asked human participants to teach the robot those names as well. In some other experiments, we asked participants to learn from the robot who acted as a teacher by uttering those object names. For another example, in adult-adult interaction, an informed confederate was asked to behave in certain ways when he was interacting with his social partner. In all of the studies, multiple sensing systems are used to simultaneously record multimodal multi-streaming behavioral data from these two agents (be they humans or robots). As shown in Figure 1, the raw data collected from various sensing systems includes:

- **Video**: there are up to 6 video streams recorded simultaneously with a frequency of 30 frames per second, and the resolution of each frame is 720x480. Approximate 180,000 image frames were recorded in a 6-minute interaction.
- **Audio**: The speech of the participants is recorded at a frequency of 44.1kHz.
- **Body motion**: there are multiple position sensors in Polhemous motion tracking system, one on each participant's body part, e.g. the head or hands. Each sensor

provided 6 dimensional (x,y,z, yaw, pitch, and roll) data points at a frequency of 120Hz. In a 6-minute interaction, we have collected 864,000 position data points from each participant.

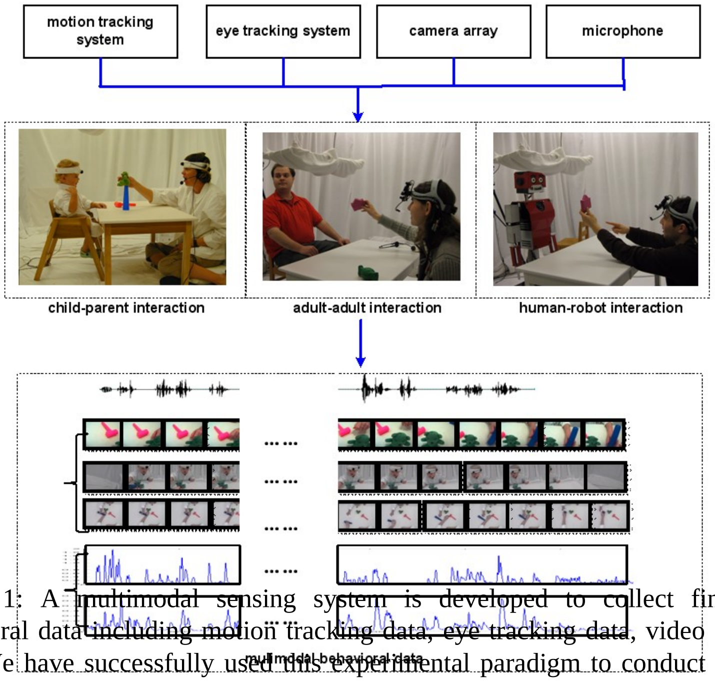

Figure 1: A multimodal sensing system is developed to collect fine-grained behavioral data including motion tracking data, eye tracking data, video and audio data. We have successfully used this experimental paradigm to conduct studies of adult-adult interaction, child-parent interaction and human-robot interaction. In each study, we have collected multi-stream multimodal data for further data analysis and knowledge discovery.

- **Eye gaze**: an eye tracker records the course of a participant's eye movements over time at 60Hz.

Those data are synchronized based on a central time-clock and therefore we can easily align them to analyze sequential patterns across data streams.

3. Coding of Multimodal Data

The next component in our data-driven paradigm is automatic data coding – deriving various time series from raw multimodal data.

Video Processing The recording rate for each camera is 30 frames per sec. The resolution of each image frame is 720*480. As shown in Figure 2, we analyze the image data in two ways: (1) At the pixel level, we use the saliency map model developed in [4] to measure which areas in an image are most salient based on motion, intensity, orientation and color cues. Itti's saliency map model applies bottom-up attention mechanisms to topographically encode for conspicuity (or ``saliency") at every location in the visual input. (2) At the object level, the goal is to automatically extract visual information, such as the locations and sizes of objects, hands, and faces, from sensory data in each camera. These are based on computer vision techniques, and include several major steps. The combination of using pre-

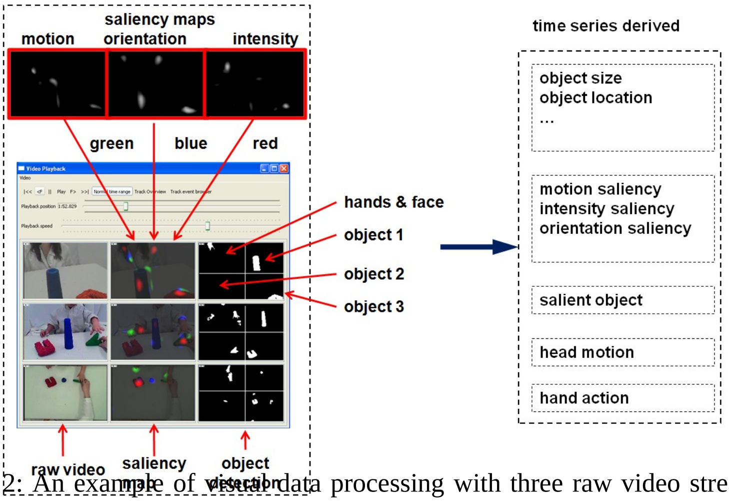

Figure 2: An example of visual data processing with three raw video streams (the left column). We calculated saliency maps from each video stream (the middle column) and also extracted visual objects from the scene (the right column). Next, a set of time series are derived based on measures such as sizes and locations of each of those objects, visual saliency and also head and hand movements.

defined simple visual objects and utilizing start-of-the-art computer vision techniques results in high accuracy in visual data processing. The technical details can be found in [5].

Motion data Processing Six motion tracking sensors on participants' heads and hands recorded 6 DOF of their head and hand movements at the frequency of 240 Hz. Given the raw motion data {x, y, z, h, p, and r} from each sensor, the primary interest in the current work is the overall dynamics of body movements. We grouped the 6 DOF data vector into position {x, y, z} and orientation data {h, p, r}, and then we developed a motion detection program that computes the magnitudes of both position movements and orientation movements. We next cluster hand movement trajectories into several action prototypes (e.g. reaching, holding, and manipulating).

Speech processing We first segment the continuous speech stream into multiple spoken utterances based on speech silence. Next, we ask human coders to listen to the recording and transcribe the speech segments. From the transcriptions, we calculate the statistics of linguistic information, such the size of vocabulary, the average number of words per spoken utterance, the frequent frames in spoken utterances.

In addition to automatic coding, we realized that some information may be hard to extract automatically. For instance, whether a person is holding an object cannot be easily detected just based on the distances between hands and the visual objects captured from video cameras as one may put hands close to an object without holding it. In light of this, we have also developed a manual coding program to allow human coders to record those derived variables. In the future, we are interested in investigating a solution that can take advantage of both fast automatic coding and potentially more accurate human coding.

Overall, as a result of sensory data processing (shown in Figure 2), we derived more than 200 time series from various sensing systems and they capture different kinds of perception and action variables in multimodal social interaction, such as sizes and saliency of objects from different cameras, trajectories of head and hand movements, and the distances of objects to a participant. This data processing route has been used in different studies we've conducted using the multimodal sensing system.

4. Knowledge Discover and Data Mining

With huge amounts of multimodal data collected from various studies of human-human and human-robot studies, a critical component in data-intensive discovery is the ability to discover new patterns from such data. Now how can we discover what we do not know what we are looking for? Classic approaches to data-mining often require that one has some idea of what one is looking for. More specifically, advances in machine learning and data mining provide tools for the discovery of only pre-defined structures in complex heterogeneous time series (hidden Markov models, dynamic time warping, Markov random fields, to name a few, e.g. [6]). A relevant research field to knowledge discovery is information visualization with the goal to visually present data to highlight certain aspects of patterns and structures so that researchers can easily spot interesting patterns based on their visual perception systems. As shown in Figure 3, data mining and information visualization are traditionally treated as two irrelevant topics. Data mining algorithms rely on mathematical and statistical techniques to discover patterns while information visualization researchers are interested in finding novel techniques to visually present data in more informative ways. However, it is worth noting that these two topics share the same ultimate research goal – building tools and developing new techniques to allow users (e.g. researchers) to obtain a better understanding of massive data.

Discovering *new* knowledge requires the ability to detect unknown, surprising, novel, and unexpected patterns. For this purpose, we propose that the techniques developed in data mining and information visualization can be integrated to build a better pattern/knowledge discovery system. In the case of our multimodal data, a huge amount of information must be cut and summarized to be useful. But statistics and measures extracted from raw data may exclude embedded patterns or even be misleading. We need a mechanism to represent the overall statistics but still make fine-grained data accessible. Information visualization provides a unique opportunity to accomplish this task. Often, potential users of information visualization are not aware of the benefits of visualization techniques on data mining; or they use those techniques only as a *first* phase in the data analysis process. As shown in Figure 3 (right), we suggest a more interactive mode between data mining and information visualization. In our system, researchers can not only visualize raw data at the beginning but also visualize processed data and results. In this way, data mining and visualization can bootstrap each other – more informative visualization based on new results will lead to the discovery of more complicated patterns which in turn can be visualized again to lead to more findings. More importantly, researchers play a critical role in this human-in-the-loop knowledge discovery by applying data mining techniques on the data, examining visualization results and deciding what to focus on at the next round based on theoretical knowledge in one's mind. In this way, domain knowledge, computational power, and visualization tools can be integrated together to allow researchers to reduce the search space created by huge datasets, and quickly and effectively identify interesting patterns. In the following, we will briefly introduce visualization and information tools we developed and used in our data-driven research paradigm.

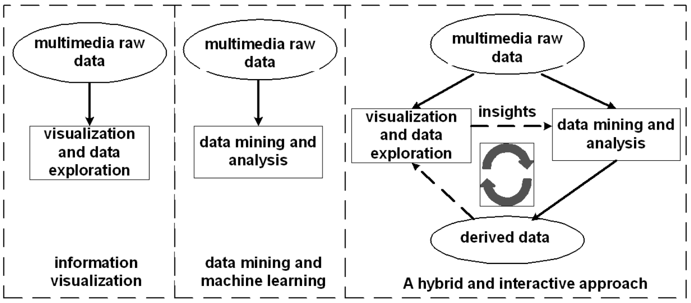
multimedia raw data
visualization and data exploration
information visualization
multimedia raw data
data mining and analysis
data mining and machine learning
multimedia raw data
visualization and data exploration
insights
data mining and analysis
derived data
A hybrid and interactive approach

Figure 3 **Left**: Information visualization techniques focus on developing informative ways to visualize data. **Middle**: Data mining algorithms rely on mathematics and statistics to find complex patterns from data. **Right**: While data mining and information visualization are traditionally treated as two separate topics, they share the same research goal. In light of this, our hybrid and interactive approach builds the links between these two and by doing so forms a closed loop between visualization and data mining.

4.1 Visual data mining system

As we discussed earlier, most data mining algorithms can effectively search and discover only pre-determined patterns and those patterns need to fit a specific definition of statistical reliability. This limitation significantly constrains what can be achieved. In light of this, we have developed a hybrid approach that allows us to use data mining as a first pass, and then from these suggested interesting/unusual patterns, we perform more directed, more detailed and deeper analyses with human inspection -- adding domain-specific expertise as a part of data analysis, followed by more automatic data mining. This idea of interactive human-in-the-loop data mining is implemented by our information visualization system [7] which has 3 key components: (1) a smooth interface between visualization and data mining; (2) a flexible tool to explore and query temporal data derived from raw multimedia data; and (3) a seamless interface between raw multimedia data and derived time series and events. We have developed various ways to visualize both temporal correlations and statistics of multiple derived variables as well as conditional and high-order statistics. Our visualization tool allows us to explore, compare, and analyze multi-stream derived variables and simultaneously switch to access raw multimedia data. As shown in Figure 4, our visualization system follows the general principles of building scientific data visualization systems: “overview, zoom & filter, details-on-command” as proposed by [8]. We embody these principles in the case of multimedia visual data mining of social and behavioral data, which comprises two major display components as shown in Figure 4: a multimedia playback window and a visual data mining window. The multimedia playback window is a media player that allows us to access and process raw multimedia data, and plays them back in various ways. The visualization window is the main tool that allows us to manipulate and visually explore derived data streams to discover new patterns. More importantly, when we visually explore a dataset, these two display windows are coordinated to allow us to switch between synchronized raw data and derived data. We have developed various functions to visualize derived data streams individually or together to highlight different aspects of multimedia multivariate data (see [7] for details).

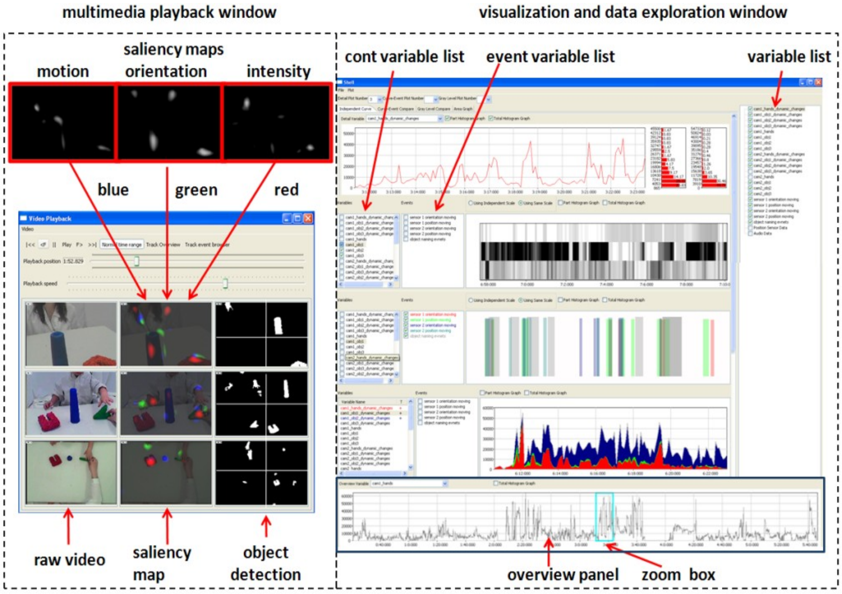

Figure 4: There are two major display components in our visualization system: a multimedia playback window (left) and a visualization window (right). The multimedia playback window is a digital media player that allows us to access video and audio data and play back both raw and preprocessed data in various ways. The visualization window is the main tool that allows us to visually explore the derived data streams and discover new patterns.

As an example to demonstrate the utilities of visual data mining, Figure 5 shows a set of eye movement data (e.g. where a person is looking at). Each stream in the figure corresponds to the same derived variable but collected from different participants. In this example, since all of the data streams are temporally aligned, we can easily compare those time series to spot interesting patterns. More specifically, there are at least two immediate outcomes from this visual exploration. First, we can discover those moments that all the participants behave similarly. This pattern can be spotted by examining multiple time series vertically. Moreover, we can visually examine the timing of their eye movement behaviors which usually cannot be captured by statistical analysis. Meanwhile, by examining those data streams horizontally, we can also easily find those individuals who are different with others (e.g. participant 3 in this example). Moreover, this explicit and informative visualization allows us to also discover *in what ways* those individuals are different. In the above example, participant 3 seems to always generate eye movements right before most of people do so – the pattern that can be easily detected through this visualization.

4.2 Information-Theoretic Measures

In multimodal communication, one agent's activities are embedded in, influenced by, and influence the momentary behaviors of the other social partner such that the whole human-human interaction can be viewed as two coupled complex systems. As the

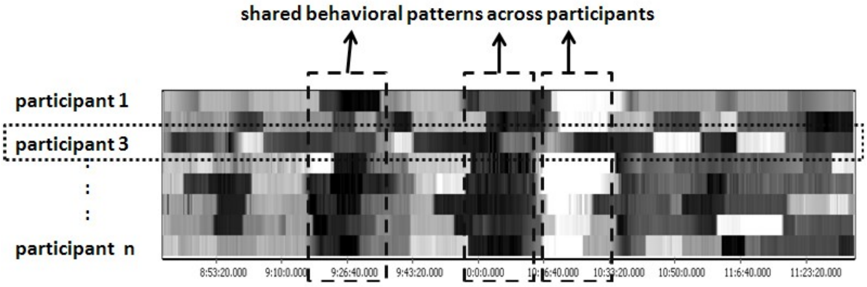

Figure 5: Application of our visualization program to the problem of comparing the same derived variable collected from multiple participants. By visually exploring those multiple streams in parallel, researchers can easily detect the shared behavioral patterns across participants. Namely, most participants generate the same behaviors or perceive the same information at those moments. At those moments, they also demonstrate individual differences. Moreover, we can also easily identify individuals that are different with the whole group. For instance, participant 3 seems to be an outlier whose gaze data are quite different with other participants.

first steps to understand those two coupled complex systems, we use information-theoretic measures to measure information flows between various time series derived from the same agent and those time series derived from two agents. We will use the data from child-parent interaction as an example here. Given multi-streaming continuous time series extracted from child-parent interaction, we are interested in quantifying this multimodal inter-person information exchange at the bit level. Our data mining and pattern discovery process consists of two steps. We need to first convert a continuous time series into a discrete stream of system states so that we can form probabilistic distributions of the states of each variable over time. We can then apply information metrics (e.g. entropy and mutual information) to quantify the amount of information in bits.

The discretization technique we employed in the first step is based on Symbolic Aggregate Approximation (SAX) [9]. The goal here is to provide an efficient and accurate symbolic representation of time series which makes it easier to apply information theoretic measures. In brief, SAX first transforms the input time series data into Piecewise Aggregate Approximation (PAA) representation and then symbolizes the distribution space of PAA representation into a set of discrete symbols. Compared with other approaches, SAX allows lower-bounding distance measures to be defined on the symbolic space that are identical with the original data space. Thus, the information loss through this symbolization and its potential effects on subsequent data processing is minimal when we convert time series into this efficient symbolic representation.

With symbolic representations of various derived time series from multimodal human-human communication, we have applied and integrated various temporal data mining algorithms in our system. Here we will use *entropy rate* [10] as a simple example, which measures uncertainty in time series when the previous state of a series is given or the previous state of two series are given. In our case, the *entropy* *rate*, which measures the uncertainty in a distribution of the transitive probability from one state to another would be a suitable choice. Suppose that $I = \{i_1, i_2, \cdots, i_N\}$ is series of symbols, and the entropy rate of the series *I* is as follows.

$$H(I) = -\sum p\left(i_{n+1}, i_n^{(k)}\right)\log p\left(i_{n+1} \mid i_n^{(k)}\right)$$

where $p\left(i_{n+1} \mid i_n^{(k)}\right) = p\left(i_{n+1} \mid i_n, i_{n-1}, \cdots, i_{n-k+1}\right)$.

We next calculated the entropy rates of individual variables extracted from raw multimedia data. To capture temporal dynamics of each variable, instead of calculating the overall entropy over time, we defined a sliding window with the size of 2 seconds and applied this moving window to time series data so that a local entropy time series was computed. Figure 6 shows several individual entropy measures in parallel. The first two entropy time series H1 and H2 are from the child while H3 and H4 are from the parent. More specifically, H1 measures the entropy of the object held by the child's hands. H2 measures the entropy of the child's hands. H3 measures the entropy of the same object from the caregiver's view (captured by the caregiver's camera) and H4 measures the entropy of the parent's hands. By aligning these individual entropy measures side by side, we can immediately detect several patterns between these variables: 1) H1 and H2 are quite correlated; 2) Moreover, H2 is a precursor of H1 since the manipulation of the object causes the changes of the visual appearance of that object from the camera's view; 3) H3 and H4 are not correlated since the object is in the child's hands but not the parent's hands; and 4) H1 and H3 are not correlated since the object manipulated by the child is barely seen from the parent's view. Moreover, we can spot potentially important temporal events. For example, the red box in Figure 6 highlights such an event. In this example, the parent briefly took the object away from the child which causes the increase of the entropy on the parent's side (both the hands and the visual object) and as well as the decrease of the entropies on the child's side (the visual object is farther away; and the child's hand is not manipulating the object). From this example, we suggest that even simple information-theoretic measures, such as the entropy rate of a temporal variable, can lead to insightful results to quantify the dynamic communication between the two agents. We have developed various other measures such as mutual information and transfer entropy to quantify information flows not only in child-parent interaction and human-robot and adult-adult interaction.

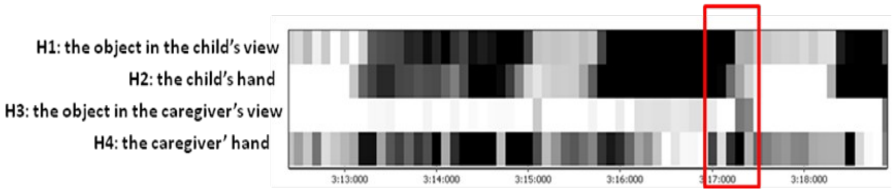

Figure 6: An example of entropy rates from four time series – two from the child and two from the parent. From this example, we can see the moment with consequential information flows between those four variables and as well as which variable leads to changes of other variables.

4.3 Interactive Data Mining

We argue that a critical challenge of data-intensive discovery lies in not only pure amounts of data that need to be mined but also the huge search space of potentially interesting patterns created by the data. For example, in our dataset, one can treat each time series independently and by doing so, we just focus on patterns within a single time series. However, if we truly view multimodal communication between two agents as two complex systems, we need to analyze correlations embedded in a subset of two, three or more time series which leads to combinational exploration. A solution for this challenge in particular and for intelligent data analysis of scientific discovery more generally clearly relies on advanced data mining algorithms to process huge amounts of data and extract potentially interesting patterns. But more importantly, we suggest that the exploratory nature of scientific discovery also requires an interactive platform allowing scientists to be in the loop of data mining, examine the current results generated and then guide data mining algorithms toward the right directions, which significantly enhances both effectiveness and efficiency of knowledge discovery. Thus, this human-guided computation can seamlessly integrate human expertise and statistical pattern discovery power. Toward this end, our data mining system supports various procedures that allow us to examine both raw and derived data, and gain insights and hypotheses about interesting patterns embedded in the data. All this is accomplished by human observer's visual system. In order to quantify and extend these observations, researchers need to develop and use data mining algorithms to extract and measure the patterns detected in visual exploration. We notice that different researchers may have different preferences of programming languages and may prefer to use certain software packages. To increase the flexibility to be compatible with data mining, our system allows users to use any programming language to obtain new results. Thus, data researchers can implement new data mining algorithms using their own analysis tools (from Matlab, to R and to C/C++) and as far as they write the results into text files (e.g. CSV form) with pre-defined formats, our system monitors user's workspace in real time and will automatically load new derived variables into the variable list so that we can immediately visually examine these new results. In this way, our visualization system supports a close and flexible coupling between visual exploration and data mining. The insights gleaned from visualization can be used to guide further data mining. Meanwhile, the results from the next round of data mining can be visualized which allows us to obtain new insights and develop more hypotheses with the data. Overall, our data analysis system is "open-minded" by not adding any constraints, assumptions or simplification on raw and derived data, but instead allows us to guide the direction and systematically explore the data through informative visualization, which is truly the power of visual data mining.

5. Conclusions

Human-human and human-robot multimodal communication can be viewed as two coupled complex systems interacting with each other through perception and action. Inspired by this conceptualization, we invented and implemented a novel data-intensive paradigm that allows us to investigate micro-level behaviors, such as head turns and gaze shifts, generated by autonomous agents (humans or robots). To achieve this goal, our paradigm includes a multimodal sensing system to collect fine-grained sensory data, automatic coding programs to extract temporal variables from raw multimedia data, a visual data mining system to visualize multi-streaming data, a set of quantitative measures (e.g. information theoretic measures) to calculate information exchanges between variables and states derived from two agents, and an interactive framework that allows us to integrate information visualization and data mining. We have implemented each of these components and have begun to use this new paradigm to discover new knowledge in both human-human communication [5] and human-robot interaction [11]. We argue that data-intensive discovery approaches have great potentials to lead to lots of breakthroughs in cognitive and behavioral studies in the near future.

Acknowledgement

This research is supported by NSF BCS 0924248 and AFOSR FA9550-09-1-0665. We would like to thank Amanda Favata and Damian Flicker for data collection, and Yiwen Zhong for contributing to software development of the information visualization program.

References

1. Bell G., Hey T., Szalay A.: Computer science. Beyond the data deluge. Science 323, 1297-1298 (2009)
2. Cohen, P.R., Adams, N.: Intelligent Data Analysis in the Twenty First Century. In Proceedings of the Intelligent Data Analysis Conference, Lyon, France, (2009)
3. Baron-Cohen, S.: Mindblindness: an essay on autism and theory of mind. MIT Press/Bradford Books (1995).
4. I tti, L, Koch, C., Niebur, E.: A Model of Saliency-Based Visual Attention for Rapid Scene Analysis IEEE Transactions on Pattern Analysis and Machine Intelligence 20(11):1254-1259 (1998)
5. Yu, C., Smith, L.B., Shen, H., Pereira, A.F., Smith, T.G.: Active Information Selection: Visual Attention Through the Hands. IEEE Transactions on Autonomous Mental Development, Vol2, 141-151, (2009)
6. Oates, T., Cohen, P.R.: Searching for Structure in Multiple Streams of Data. In Proceedings of the Thirteenth International Conference on Machine Learning, pages 346 – 354 (1996).
7. Yu, C., Zhong, Y., Smith, T., Park, I., Huang, W.: Visual Data Mining of Multimedia Data for Social and Behavioral Studies. Information Visualization vol. 8, 56-70 (2009).
9. Lin, J., Keogh, E., Li, W., Lonardi, S.: Experiencing SAX: A Novel Symbolic Representation of Time Series. Data Mining and Knowledge Discovery Journal. p. 107-144, (2007).
10. Kantz, H. and Schreiber, T.: Nonlinear Time Series Analysis., Cambridge University Press, Cambridge (1997).
11. Yu, C., Scheutz, M., & Schermerhorn, P.: Investigating Multimodal Real-Time Patterns of Joint Attention in an HRI Word Learning Task. 5th ACM/IEEE International Conference on Human-Robot Interaction, (2010).