JOLTIDA

JOURNAL OF LEARNING AND TEACHING IN DIGITAL AGE

Review Paper

Data Science and Human Behaviour Interpretation and Transformation

Ajit Singh *a

a(ORCID ID: 0000-0002-6093-3457), Patna Women's College, Bihar, India, ajit_singh24@yahoo.com
*Corresponding author

ARTICLE INFO	ABSTRACT
	The purpose of this paper is to analyze various dimensions for measurement of human behavior. Human behaviour is complex. Behaviors, emotions, cognitions, and attitudes can rarely be described in terms of one or two variables. It is multimodal in nature. Furthermore, the traits, modalities and dimensions cannot be measured directly, but must be inferred from constructs which in turn are measured by multiple factors or variables. I have emphasized on the use of baseline data for each subject as the degree of expressiveness for same situation which varies for each subject and needs to be measured based on the individual trait of the subject. This can be done by making baseline data for subjects being researched. Subsequently, discussion has been done on data analysis. Finally, framework for the same has been proposed. Basically, the researcher asks two questions, "Do I have anything important?" (Which is based upon the researcher's observations of some aspect of human behavior adequately addresses the observation) "If so, what do I have?" (What is the best explanation of the relationship between the variables?)

INTRODUCTION

Human behaviour is complex. Behaviors, emotions, cognitions, and attitudes can rarely be described in terms of one or two variables. It is multi-modal in nature. Furthermore, the traits, modalities and dimensions cannot be measured directly, but must be inferred from constructs which in turn are measured by multiple factors or variables. The researcher must construct his/her view of reality (observations about humans), create multiple measures to assess the constructs, and conduct an appropriate statistical analysis of the data, and correctly interpret the outcomes (Ekman, 2012). Basically, the researcher asks two questions,

"Do I have anything important?" (Which is based upon the researcher's observations of some aspect of human behavior adequately addresses the observation)

"If so, what do I have?" (What is the best explanation of the relationship between the variables?)

The learning for the research problem is that the human behavior is multimodal. The content comes from various inputs like face expressions, gestures captured from images, video, audio, text etc i.e. multimedia. There is possibility of using XML for defining CO (Content Objects) emerging from the multimodal nature of human behavior and allow creation of multi-layered structure, integrating intrinsic or latent properties of the content, dynamic properties, non-verbal expressions, emotional as well as real-world descriptors.

System Conditioning is adoption of required devices to capture source information based on application. In case the application is behavior interpretation of students the conditioning means database of students, teachers, setting devices in classroom, laboratories, creation of individual databases defining aspects like personality, history of academic achievements etc. In case application is related to office environment the experience record, skill sets data, setting of data capture in office environment. The conditioning depends on application planned. Similarly the comparison matrix needs to be defined. The comparison matrix defines the expected behavior and helps in measurement of intensity of the behavior measured.

Subject Dimensions in Human Behavior

The subject dimension in human behaviour are as followings;

- Historic data (background, exposure, experience, achievement till date)…portfolio self/peer evaluated/annotated by experts
- Posed Cues (creation of baseline data images, videos, speech samples, biometric data)… creation of baseline data giving extremities of behavior expression example neural face…sleepy/alert face. This would set a scale for each individual which would help to measure degree of expression of a behavior)
- "Ekman" (Ekman, 2012) has defined six basic states in the FACS. Further, researchers have restricted these to four
- Current State of subject (internal motivation, Mood, immediate events in life of the subject, ability/ competence for current challenges)

- Capturing the habits is important to associate it with the subject
- Personality, Voice, Written data for baseline
- Psychologists also mention self-esteem, confidence

I have emphasized on the use of baseline data for each subject as the degree of expressiveness for same situation varies for each subject and needs to be measured based on the individual trait of the subject. This can be done by making baseline data for subjects being researched.

Situational Dimensions in Human Behavior

The situational dimension in human behaviour are as followings;

- The event / situation description: This is important to know whether situation where the behavior measurement is being done is formal (classroom, workplace etc.) or informal (home, friend circle, party), emergency (hospital, disaster) etc.
- Place, Time, Environment
- Challenges posed by a situation (game final stages, exam, expectations from organization, interview)
- ABC Model of attitudes (Fogg, 2010) has involved the dimensions in a situation based on (dominance submissiveness, arousal-non-arousal, pleasure- displeasure proposed by Mehrabian and Russell developed in 1974 to ABC Model of Attitudes)
- Measurements have been done on valence, potency, arousal states
- Fogg (2010) explains behavior as B = Ability * Motivation * trigger

Depending on the situation, the subjects will exhibit different behavior; hence the situation helps in setting the threshold for the expected behavior. It would help in comparing desired threshold with exhibited thresholds of the subjects and may be helpful in setting behavior transformation strategies.

Other Human Dimensions

Since human behavior is reflection of social interactions the characteristics of the other players who influence the behavior is also important;

- Dominance-Submissiveness in subject- peer interactions
- Motivation in situation
- Personality, Familiarity, Image of external players

Human behavior is indeed influenced by the surrounding subjects and the surrounding subjects determine the exhibited behavior of the subject in given situation.

Objective Dimensions

Spontaneous Cues (data captured in a situation like classroom, work place, skill demonstration) would be recognized, improvised and feature extraction would be done and classification for identification of behavior from

- Capturing of Macro expressions (to capture situational behavior personality, voice of the subject in current situation)
- Capturing of micro expressions (initiation of expression, reaching peak and fading of the expression)

Thus, the Data on Human Behavior would essentially include:

- Subject Dimensions
- Situational Dimensions
- Other humans Dimensions
- Objective Dimensions

Understanding the Data Science

What is the data science?
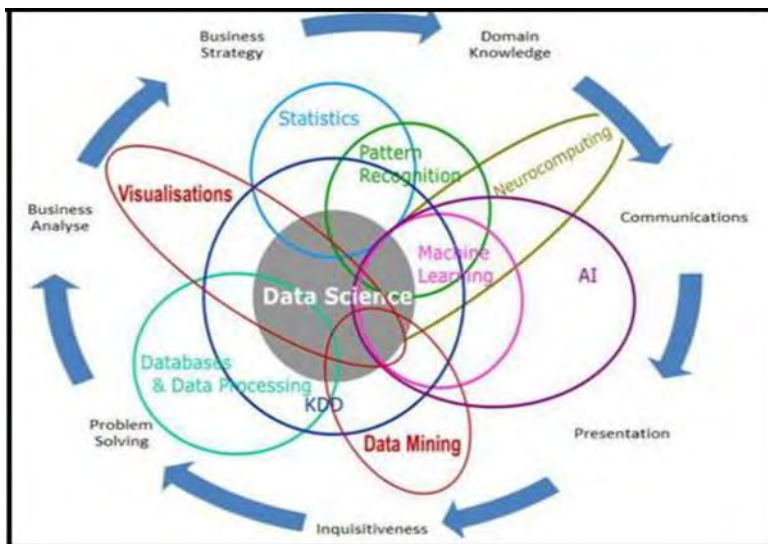
Figure 1. Interdisciplinary nature of Data Sciences. Adapted from Managing Therapy-Interfering Behavior: Strategies From Dialectical Behavior Therapy (p. 172), by A. L. Chapman and M. Z. Rosenthal, 2016, Washington, DC: American Psychological Association. Copyright 2016 by the American Psychological Association.

The Figure 1 shows the universe of data sciences that is interdisciplinary in nature and includes;

- Statistics (Statistics is the study of the collection, analysis, interpretation, presentation, and organization of data. In applying statistics to, e.g., a scientific, industrial, or societal problem, it is conventional to begin with a statistical population or a statistical model process to be studied)
- Pattern Recognition (Pattern recognition is the process of classifying input data into objects or classes based on key features. There are two classification methods in pattern recognition: supervised and unsupervised classification.)
- Artificial Intelligence (Artificial intelligence (AI) is the intelligence exhibited by machines or software. It is also the name of the academic field of study which studies how to create computers and computer software that are capable of intelligent behavior)
- Neuro-computing (a computer or a software program that uses a neural network simulating the human brain and can be trained to perform specific tasks, as pattern recognition)
- Machine learning (Machine learning is a subfield of computer science that evolved from the study of pattern recognition and computational learning theory in artificial intelligence. In 1959, Arthur Samuel defined machine learning as a "Field of study that gives computers the ability to learn without being explicitly programmed")
- Data mining (the practice of examining large pre-existing databases in order to generate new information.)
- Knowledge Discovery from Data (Knowledge discovery in databases (KDD) is the process of discovering useful knowledge from a collection of data. This widely used data mining technique is a process that includes data preparation and selection, data cleansing, incorporating prior knowledge on data sets and interpreting accurate solutions from the observed results.)
- Databases and Data processing (the carrying out of operations on data, especially by a computer, to retrieve, transform, or classify information.)
- Visualization (Visualization is the process of representing abstract business or scientific data as images that can aid in understanding the meaning of the data)

In the problem of human behavior interpretation, there would be need of the tapping the different data science dimensions.

Data Science is interdisciplinary in nature and needs skill sets and Figure 2 shows the skill sets required for Data Science which includes;

- Hacking skills (to acquire, clean and manipulate the massive data)
- Math and Statistical knowledge (to choose appropriate method and tools to extract insight from data)
- Substantive Expertise for generating motivating questions, hypothesis and interpreting results

The Figure 2 also shows the interesting zones. The most critical zone to be avoided is the Danger Zone where hacking skills are combined with substantive expertise without rigorous methods which can lead to incorrect analysis. Thus, it is important to address the problem of human behavior interpretation with the above three skills.
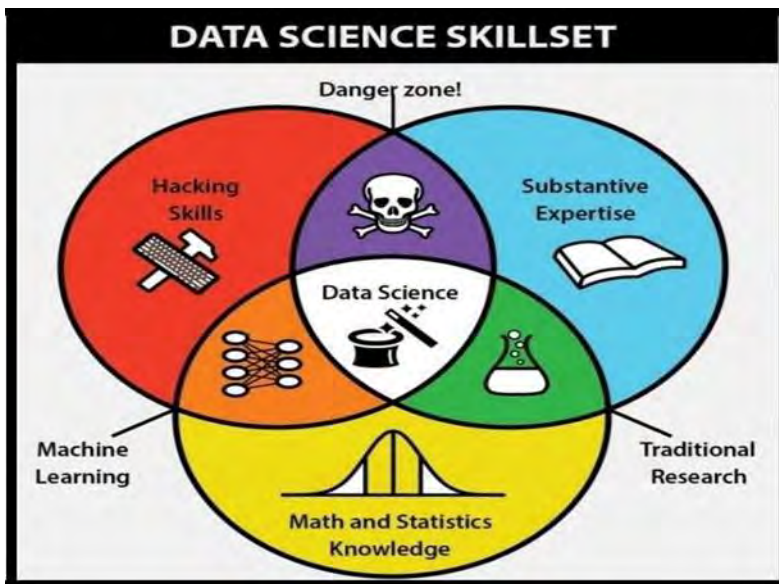
Figure 2. Data Science Skill sets. Adapted from Managing Therapy-Interfering Behavior: Strategies From Dialectical Behavior Therapy (p. 119), by A. L. Chapman and M. Z. Rosenthal, 2016, Washington, DC: American Psychological Association. Copyright 2016 by the American Psychological Association.

The Workflow involved in data sciences has been shown in Figure 3 which involves;

- Data acquisition and cleanup
- Storage and management
- Analysis
- Visualization
- Communication
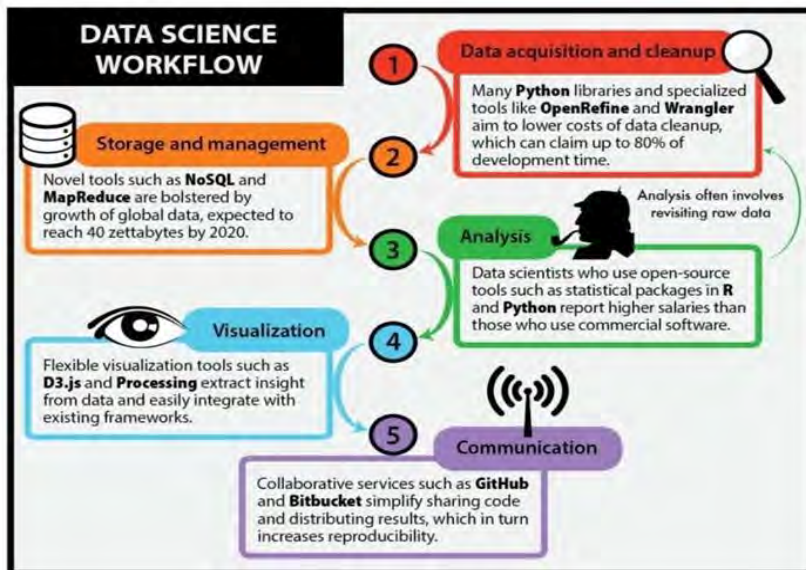
Figure 3. Data Science Workflow. Adapted from Managing Therapy-Interfering Behavior: Strategies From Dialectical Behavior Therapy (p. 63), by A. L. Chapman and M. Z. Rosenthal, 2016, Washington, DC: American Psychological Association. Copyright 2016 by the American Psychological Association.

Proposed Framework for Data Science & Human Behavior Analytics

The Proposed Framework for Data Science & Human Behavior Analytics are as followings;

- Automatic classification of human behavior involves
- Understanding of bodily motion,
- Understanding of gestures and signs,
- Analysis of facial expressions,

- Interpretation of affective signals
- Social signal processing deals with interactions between humans integrating verbal cues with non-verbal behavioral cues

'Zahariadis' (Zeng, 2009) have stated about authoring of content objects using XML. The Content Objects (COs) are Objects that have representation of a specific instance of either a physical object or a physical entity (an entity that has physical existence, e.g. an human being) or an abstraction (a general concept formed by extracting common features e.g : learning engagement, work efficiency),an event or a concept (human beings learning, humans at work, humans moving on street), which might have multiple views (many images, videos, audio files, text i.e multimedia ; real-world and user-related information i.e context, characteristics of the subject). Content Objects thus have a polymorphic container, which may have media, rules, behavior, relations and characteristics or any combination of the above.

The Authoring Tool have to define the content object with different types of media items, real-world information (location, context, weather, time, etc.) and user-related information (emotional, expressive characteristics, prior experience, prior learning etc.) and produce a rich media representation which is a Content Object (P. Daras et al., 2010) have described RUCoD (Rich Unified Content Description in their paper " Introducing a Unified Framework for Content Object Description" and "The formal description of a CO is its RUCoD file, which is an XML-based document specifying descriptors for all the above input types. 'Vincenzo Croce' have described the RUCoD content object descriptor initially designed to serve needs of I-Search Framework and extended for Project CUBRIK.

The learning for the research problem is that the human behavior is multimodal. The content comes from various inputs like face expressions, gestures captured from images, video, audio, text etc i.e. multimedia. There is possibility of using XML for defining CO (Content Objects) emerging from the multimodal nature of human behavior and allow creation of multi-layered structure, integrating intrinsic or latent properties of the content, dynamic properties, non-verbal expressions, emotional as well as real-world descriptors.

The proposed model is shown in Figure 4. There are three stages:

- System Conditioning
- System Operations
- System Transforming
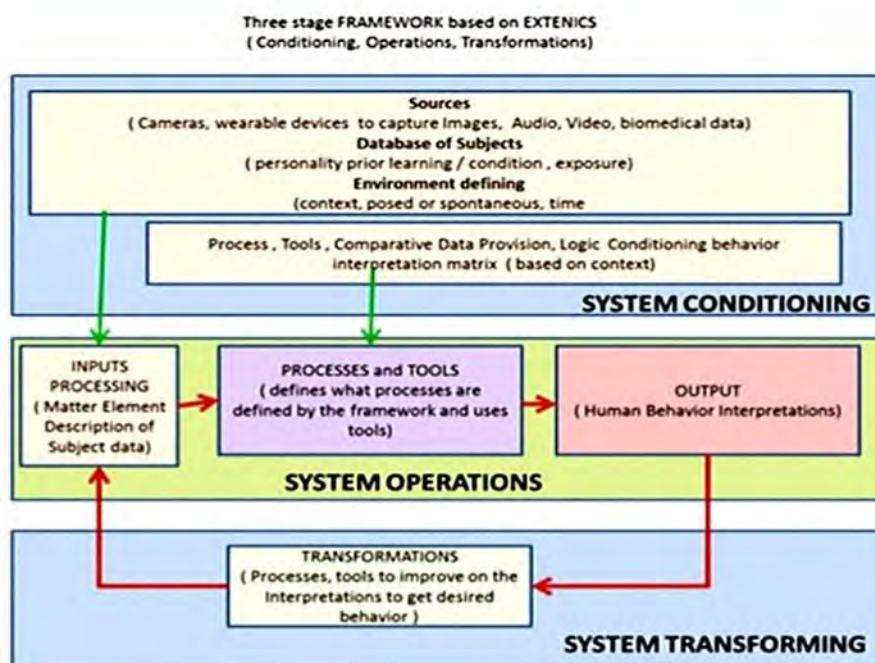
Figure 4. Framework for Human Behaviour Interpretation and Transformation with Data Science. Adapted from Managing Therapy-Interfering Behavior: Strategies From Dialectical Behavior Therapy (p. 283), by A. L. Chapman and M. Z. Rosenthal, 2016, Washington, DC: American Psychological Association. Copyright 2016 by the American Psychological Association.

- System conditioning: This stage is essentially for creation of subject data, improving the captured data (face enhancement, 3D modeling to get good input images for analysis, tabulate captured data for subjects involved)
  - o Mapping of subject data Background Experience, cultural background, financial background, friend circle, Social media cues taken from face book, twitter and other content
  - o Further earlier work profile record, academic performance analysis, personality analysis, counseling reports, health reports all contribute to describe the subjects

o   System Conditioning is adoption of required devices to capture source information based on application. In case the application is behavior interpretation of students the conditioning means database of students, teachers, setting devices in classroom, laboratories, creation of individual databases defining aspects like personality, history of academic achievements etc. In case application is related to office environment the experience record, skill sets data, setting of data capture in office environment. The conditioning depends on application planned. Similarly the comparison matrix needs to be defined. The comparison matrix defines the expected behavior and helps in measurement of intensity of the behavior measured.

The system conditioning thus gives a refined input to the next stage to eliminate system errors due to unconditioned data. The Context or Situation is the determining factor of behavior measurement. The context can be teacher student interaction in classroom, boss, peers in office, team of friends in a relaxed environment etc. The context determines the comparison matrix for behavior interpretation.

System Operations include input, processes and output:

*   Inputs: The inputs for human behavior interpretation framework could be Images, Video, and Audio Data for the context and environment, prior historical data of the subject. The data capture can be for individual subjects or even for a group as interactions happen in groups and the source, channel and receiver.
*   Process and Tools: The human behavior interpretation is multimodal in character and needs to be processed accordingly. Since several tools, methodologies are available. Appropriate ones can be used; new processes can be evolved using various theories. Authors are using Extenics Theory for Evolution of framework
*   Output: The output is the interpretation based on inputs and the process, tools used.

The interpretations give compliance or divergence from to expected behavior. Based on the degree of divergence behavior interpretation is proposed.

System Transformations

Extenics Theory has demonstrated that matter can be changed as desired. The subjects are given the behavior change / transformation opportunity through different interventions. These are interventions done for systems environments where a desired behavior is required and the measurements are again done to ensure desired behavior from the subjects. The process would be iterative in nature and behavior desired would be function of time, transformations and challenge opportunities for the subjects. The measurements are again done by System Operations stage and degree of variation in term of divergence is indicative of behavior change due to transformation strategy given.

The use of Extenics adds new dimension of transformation thereby making the model more holistic.

The above model involves various aspects of data sciences and also proposes possibility of behavior transformation.

Possible Application areas for Framework

There are several areas for application that are as follows:

*   Academic Institutions to measure effectiveness of teaching learning
*   Work places to monitor performance of tasks and identify possible ways of improvement
*   Work places to analyse aspects to increase productivity
*   Tracking of patients in Healthcare centers
*   Observing civic habits of people and transforming

CONCLUSION

The identified problem of "Human Behaviour data science for analytics" has multiple dimensions starting from understanding possible dimensions of human behavior. The authors have identified the dimensions which are holistic with provision to expand. The data sciences is an important element for working upon the data created by subjects to interpret, identify behavior. Extenics Theory has demonstrated that matter can be changed as desired. The subjects are given the behavior change / transformation opportunity through different interventions. These are interventions done for systems environments where a desired behavior is required and the measurements are again done to ensure desired behavior from the subjects.

The process would be iterative in nature and behavior desired would be function of time, transformations and challenge opportunities for the subjects. The measurements are again done by System Operations stage and degree of variation in term of divergence is indicative of behavior change due to transformation strategy given.

The framework proposed is holistic encompassing the human behavior dimensions and the data sciences dimensions making it authentic.

Ethics and Consent:

Ethics committee approval is not required as it does not involve clinical researches on humans as well as it does not contain Retrospective studies in accordance with the Law on Protection of Personal Data.

REFERENCES

Ekman, S. (2012). *Data Science Insights*. Retrieved from https://www.paulekman.com/

Iris, B. (2014, Aug 12). *Pleasure, Arousal, Dominance and Russell revisited*. Retrieved from http://link.springer.com/article/10.1007/s12144-014-9219-4

Fogg, R. (2010). Human Behavior Interpretation, Retrieved from https://www.BehaviorGrid.org

Moeslund, T. (2016, Feb 22). A survey of advances in vision-based human motion capture and analysis. *Computer Vision and Image Understanding, 104*(2-3), 90–126.

Sebe, N. (2012). Communication and automatic interpretation of affect from facial expressions. *Affective computing and interaction: psychological, cognitive, and neuroscientific perspectives, 8*(2), 114-115

Vinciarelli, J. (2014). Social signal processing: Survey of an emerging domain. *Image and Vision Computing, 27*(12), 1743–1759.

Wang, L. (2003). Recent developments in human motion analysis. *Pattern Recognition*, *36*(3), 585–601

Zeng, Z. (2009). A survey of affect recognition methods: Audio, visual, and spontaneous expressions. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 31*(1), 39–58