Behavioral and Perceptual Models for Secure Data Analysis and Management

Sandesh Achar

Staff Engineer, Intuit Inc., Mountain View, California, USA

*Corresponding Contact:
Email: sandeshachar26@gmail.com

Manuscript Received: 25 Sept 2019 - Revised: 15 Dec 2019 - Accepted: 29 Dec 2019

ABSTRACT

The ability to monitor and forecast citizen behavior on a large scale developed to be a top target for governments subject to security and intelligence that are collective in the current worldwide culture where the web has become the primary medium for commerce and communication. Meanwhile, significant privacy-related issues have surfaced considering the innovative opportunities that artificial intelligence (AI) generates for collective behavior analysis when presented to governments as a way in which the government will comprehend. In the current study, we conducted an extensive literature analysis using techniques such as data mining and interviews such as in-depth to determine the primary uses of Artificial Intelligence that governments use and describe citizens' privacy issues. Our findings showed that the government employed 11 AI initiatives to enhance interactions with residents, local organizations, services offered by government agencies, and the economy, among other things. Issues are identified relating to the risk of behavior modification, intelligent decision-making, data privacy regulation and law, digital surveillance, and decision automation as they pertain to people's privacy when governments deploy AI. Finally, the report concluded that the debate of developing rules centered on the moral citizen data architecture gathering, with consequences for governments aiming to regulate security, morality, and data privacy. We also suggest a study schedule with 16 research questions to be investigated in future studies.

Key Words: Artificial intelligence, Collective Behavior Analysis, Data Model

INTRODUCTION

The evolution of artificial intelligence (AI) in recent years has prompted organizational model adaption in private and governmental organizations. Judgments and data-oriented analysis of behavior, as well as the economic global system and rules, have become crucial for public actors in the current global society, where the primary communication tool has been proven to be the internet itself (Zhuoxuan et al., 2015). The definition, conceptualization, and construction of legal and theoretical standards that might impose moral and practical boundaries on the handling and incorporation of data related to citizens, have emerged to be an obstacle in professional, legal, and academic configurations in this context-linked society.

Concerns about user privacy have arisen due to extensive research, particularly when establishing criteria for governments to choose how.

Several privacy-related issues have arisen, especially when governments decide on methods to employ AI to comprehend social behaviors, forecast their activities and movements, and take appropriate action. It should be noted that the term "AI" defines the emulation of the creation of algorithmic models about and association with human intelligence that automatically operates and learn based on inputs created by people (Achar, 2016). The incorporation of Artificial Intelligence by businesses and governments has increased exponentially in recent years, and this growth has been predicted by several of its advantages, including anal financial risks and ratio analysis, increased profitability, prognostication of complex associations, pattern and trend identification and high accurate predictions of massive data amounts (Chimakurthi, 2017).

The idea of data sciences concerning behavior has been established to various conjugate topics linked to behavior and data science, in collateral to the growth in the government's use of Artificial Intelligence and because of behavioral and data evolution (Barrionuevo et al., 2018). Several earlier studies have specified the incoming rules for its development, even though the expression "behavioral data sciences" is lacking in the scientific literature. To AI to predict human behavior, the term "BDS" is oriented to a recent and developing discipline that collaborates methods from the behavior-related humanities and sciences. Algorithms that operate with AI could be utilized by governments and systems that monitor behavior, detect trends, and study societal knowledge as well as its users or consumers by using BDS approaches. The analysis of government-developed AI strategies is directly tied to applying the BDS idea in this study. However, the present research is groundbreaking and unique because the phrase does not occur in the published scientific literature. Additionally, businesses use customer and user data to enhance their goods and services and communicate information with other parties, including interested parties like the government or other public institutions.

Therefore, the research topics that are addressed to exhaust the literature gap are:

- To forecast the behavior of society, what Artificial Intelligence tools can governments use?
- What types of privacy concerns for citizens should be anticipated when governments implement behaviorally-based AI into their plans?
- To identify definitional viewpoints of privacy issues in behavioral data science deployment in government.

Our current state of living is the most profound shift in the information age—specifically, in an environment where the primary information source is data. Therefore, it is essential to understand behavioral data sciences and surveillance capitalism. This section proves the identification of the primary theoretical stances utilized in the literature to examine the elements influencing the growth of Artificial Intelligence in governments, intending to lay out an academic history with the fundamental hypothesis employed to predict the behavior of users by analysis in the digital environment. For their part, governments must stay current and utilize cutting-edge technology to comprehend society's needs. Unfortunately, numerous projects contend that the Internet's regulation is ineffective, and the community is resilient in the insistence that data be kept a mystery at all costs.

User privacy is an issue considering this. However, users know that governments can transform their basis relating to behavior data by analyzing human encounters and steps into intelligent robots that can anticipate anything they are addressing (Achar, 2018). Therefore, understanding the behavior of the future society that governments govern will always be the goal.

BEHAVIORAL DATA SCIENCES AND SURVEILLANCE CAPITALISM

The idea of surveillance capitalism is formed out of this privacy paradigm, worries about Artificial Intelligence and its use by governments to watch, listen actively to, track potential alarm states, or predict various age occurrences that negatively affect society (Pham et al., 2017). The idea of surveillance capitalism promotes human experience as a single source of automated data for predicting human behavior.

Figure 1: Surveillance Capitalism

The idea of Capitalism of surveillance also suggests that user behavior data is sacrificed in favor of a profitable economic business to use individuals as products of enormous data production. On the other hand, there are targets of improving service and comprehending the society's culture to enhance the government's public offer.

Users' data gathered while using connected devices is the primary data source in surveillance capitalism. All this data is analyzed utilizing BDS, which adopts a novel analytical perspective by fusing many study domains. The variety of techniques governments and businesses use to acquire data has significantly expanded in recent years (Gao et al., 2014). Several variables point to metrics for gauging the behavior of users on the web or via linked devices and mobile.

The primary sources of data up until this point have included, among others, customer relationship management (CRM) systems, marketing automation sources, intelligent organization systems, websites, and mobile devices (Emani & Yamin, 2012). When this data type is used, it automatically creates groups known as targets to explain certain qualities specified by the data analysis system's organizational framework. However, as was already

mentioned, the quantity and variety of connected devices, including the Internet of Things and innovative city services, among others, —have recently multiplied tremendously.

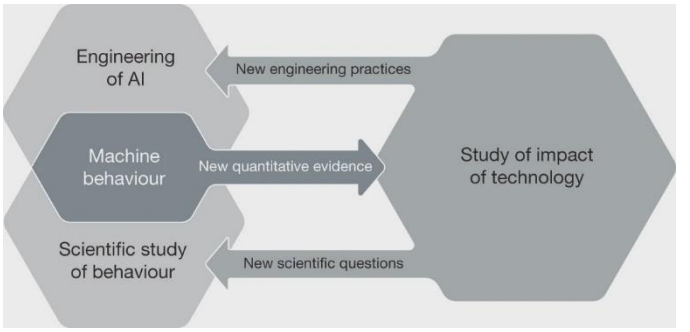

Figure 2: Behavioral Analysis and Machine Learning

Concerning these ideas, the following diagram outlines the key concepts in BDS analysis that governments can employ to track user behavior using the data they produce.

The value of data sources within an ecosystem that propels economic growth is based on behavior analysis, and its data cannot be overstated (Gao et al., 2015). However, as noted in numerous earlier studies, ethics is neither a necessary model of a business component based on vast data gathering and analysis because the awareness that users will utilize data is not known for a financial contribution by selling to interested parties. Additionally, several other initiatives may be used to safeguard people from wrongful uses of their data, such as the legislation of GDPR that the European Union and European Commission have both introduced. Although the new rule requires businesses to state how the data are used, in practice expressly, individuals need more expertise to comprehend their mobile devices and any other linked devices, the legal and privacy policies, and applications legal notices.

There is evidence that, while being aware of privacy concerns, the digital natives, Millennials, and others favor utilizing programs as quickly as possible rather than taking the time to comprehend modes in which data will be deployed to the phenomena known as "instantaneous reward." The problem of user privacy has caused considerable worry in several circumstances, including the current Covid-19 pandemic. For example, governments have chosen to invite individuals to utilize location-tracking applications throughout the crisis of Covid-19 to trace the virus illnesses and alert the public on their chance of having come into proximity with the infected. According to a late examination of these innovative monitoring methods used by citizens, governments routinely engage in this kind of active listening. The data sources that governments might have the capability to multiple firms gathering information from consumers, which has been investigated in the past through several projects. One of these studies, published by The New York Times, highlighted many of the worries governments might have when making decisions. Governments then utilize user data to enhance their procedures.

According to a review of governments' cutting-edge monitoring methods, citizens routinely actively listen. Several studies have already investigated the data sources to which governments may have access (Pahl & Xiong, 2013; Talukder et al., 2010; Zentall et al., 2002). When collecting user data from numerous corporations, governments may be concerned about several factors. Governments then prioritize the issue of national security by using user data to enhance their methods for observing society and its behavior. In this kind of strategy,

government agencies leverage user behavior data as a source to develop machine learning algorithms. Therefore, behavioral data analysis is exceptionally vital for government AI programs.

METHODS

Final contributions to the current study were chosen via a review procedure that emphasized determining the primary goals of each potentially pertinent study. Notably, the results of government research utilizing Artificial Intelligence were evaluated from the standpoint of BDS as a developing phenomenon. As a result, the "BDS" term was not employed during the operation.

The first step was discovering vague and exclusionary terms relating to the study's goals by analyzing the article's keywords and abstracts. Second, a thorough analysis of the papers that were determined to be appropriate was carried out. The next step was deciding whether the study's goals were directly or indirectly related to the current research. The topic's relevance to the study objectives was then evaluated. We also determined whether the methodology's quality and the findings' evaluation were acceptable. Finally, articles that could have defined concepts correctly concerning the goals of the current study were eliminated.

Concerning the methodological guidelines stated, we constructed the second component of the suggested strategy after conducting a thorough assessment of the literature to confirm the accuracy of the theoretical foundations. Once the government's use of AI and BDS was established as relevant, we planned and prepared for the interviews about the findings of the comprehensive literature review.

Use of Interviews

In-depth interviews were conducted with government working informants to learn more about how governments are using AI and the privacy issues they have. For example, instead of quantitatively evaluating the topic under study, these interviews sought to thoroughly grasp it by gathering data from original primary sources.

We used digital tagging in each case to facilitate further analysis using the framework of Natural Language Processing (NLP). Of the informants, some were government-employed, others were government economy specialists, and two were affiliated with groups that provide advice to the government.

Each video conference interview and the in-person interview lasted for roughly 35 minutes. Telephone interviews typically lasted 20 minutes. Responses to email interviews were usually 750–600700 words long interview data were gathered, and questions were displayed. The informants were chosen according to their current or previous work done in the government. All the sources had connections to government advisors, political parties, and public administrations. The use of Open-ended questions was employed in our semi-structured interviews.

RESULTS

Identification of the primary applications of AI used by governments in the dataset studies that have emanated from the systematic literature research findings. Based on the outcomes of this methodological procedure, the interviews were created in this manner. In addition, as previously mentioned, informant interviews were conducted with informants who currently perform in or have previously worked for governments to supplement our systematic

literature analysis findings. The critical ideas identified in the literature about users' privacy and governments' usage of Artificial Intelligence served as the foundation for the interviews. Thus, the discussions aimed to learn about novel applications of AI while simultaneously learning about user privacy and how user information is handled based on the outcomes.

The primary use of AI that governments have identified is the ongoing creation of new models that improve the effectiveness of the outcomes. This is a feature of AI because if the goals are centered on profitability, the more machine learning models are taught, the more effective they get in predicting financial outcomes. In addition, the process improvement through the decision and the development of governance and management processes were impressive.

This approach employs methods to comprehend and improve relationships with individuals through channels like social networks, information systems, or platforms for exchanging data. The findings of our research on the concepts and privacy issues uncovered in the literature review are shown in Table 7 and emphasize how simple it is for governments to access citizen data for the sake of building AI models. Specifically, public institutions work to address this reality through good governance programs. However, the right of individuals to privacy is intimately connected to legal data use and provides access to personal Intel about citizens. Using various data science techniques incorporating AI to deploy predictions, one can use data produced to anticipate people's behavior. In terms of health, society and the economy are very simple (Dhall et al., 2016).

In the current study, we looked at the primary applications and methods of artificial intelligence created by governments and the primary issues surrounding user privacy. In addition, numerous data-mining approaches were used in our analysis of the qualitative interviews, and the results produced several significant findings.

Generally, governments believe understanding citizen behavior is essential to achieving good governance. However, the predictions and correlations found in the collective behavior analysis offer substantial concerns about user privacy, as shown in prior studies on human behavior using AI approaches. The literature review also gave a complete grasp of the primary research conducted in these disciplines, identifying 11 privacy-related issues and 13 uses of AI in government. Though behavioral responses can be usefully used to anticipate potential alarm states, it is necessary to investigate how to mitigate them. For instance, network security attacks could endanger users' privacy; similarly, it will be essential to explore the privacy limits to identify the action of the population.

WORKING TECHNIQUE DESCRIPTION

The description of the technology that will be deployed in the process of ensuring the models for Behavioral and Perceptual patterns work will be a classification model (Yang et al., 2018). A short illustration of how the model will work is by classifying the different behavioral and perceptual patterns that may exist or be used by the public or citizens. Then, the behavioral and perceptual patterns will be issued as a data type, such as a CSV file containing the different perception and behavioral patterns.

The behavioral models will be defined as the results that will be obtained as a way that human beings get to interact with stimuli from the ecosystem and environment. Thus, behavioral science can be used by machine learning tools to explain how the variance in a given data set can propose and produce a model that will be able in decision-making activities revolving around behavioral and perception model making.

The techniques of this algorithm will be based on domain uncertainty where a random number of individuals will be selected, say 500 individuals are to make around 2000 decisions. The choices and decisions made have randomized parameters that will be comparable to both machine learning models as well as standardized economic models. The mode of operation where the domain uncertainty is used will be more valuable and advantageous since life and day-to-day situations revolve around random decision-making strategies made by human intelligence.

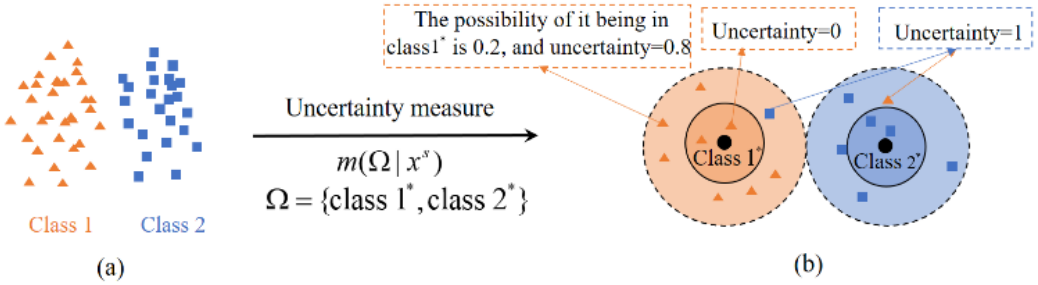

Figure 3: Domain Uncertainty

Machine learning models' primary purpose is to explain the variance that may occur when a data set poses too many unexplainable disagreements (Achar, 2018). For instance, the accuracy measuring tool in machine learning can be used to predict how the model has performed and to check the validity of the model where a higher accuracy proves that the model is better. However, in some instances, high precision, such as 100%, makes the model invalid for use because of overfitting.

EXPERIMENTAL DESIGN

The experimental design consists of randomly picking balls where each ball had a relatively equivalent price after picking it up. The total number of balls was 100, and three types of balls were comparable in color. Some balls were green, others blue, while others were red, all of which had relative prices if picked.

	Red	Blue	Green
# Balls	25	14	61
Prize	$10	$2	$0

Figure 4: Balls and their Prices

After choices were made, removing participants who did not perform the game correctly was done, leaving the number of participants to 310 after 19 options were null and spoiled. Large sets of potential lotteries were generated. This was achieved by randomizing the features {p_{red}, pg_{reen}, p_{blue}, money $_{red}$, money $_{green}$, money $_{blue}$}.

The data were divided into a test set with three questions per person and a randomly chosen training set with seven questions per person. Our core study uses the training set to calibrate various individual decision-making models. The test set is used to assess the validity of model performance when making predictions based on novel options. The metric that was employed was mean squared error.

The model that was used under choice risk was expected utility where or absolute risk aversion whereby the utility of a model was given by:

$EU(L) = p_{red}\ (money_{ed})\ \alpha + p_{blue}\ (money_{blue})\ \alpha.$

Where α is the risk aversion coefficient. The expected utility model assumes that the probabilities enter linearly into the utility.

Machine Learning Methods

The machine learning model used was regularized regression optimization, a problem model. This model produces and generates business; thus, the error caused by this reduces the chances of overfitting and lowers the mean squared error. The regression model working description dictates that the model uses a wide variety of features that can be used for prediction. The elements will be the probabilities for this scenario due to the balls.

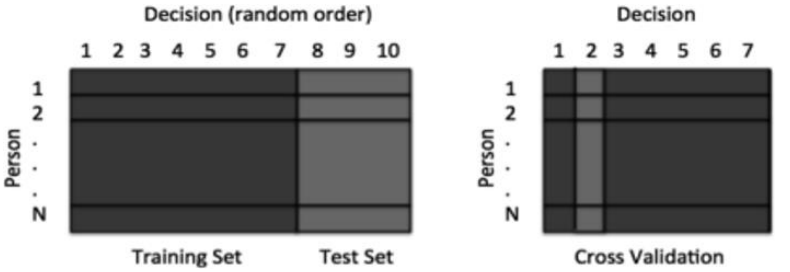

Figure 5: Using cross-validation to retrain the data

Calibration parameters will target values using the remaining data by delaying three decisions for each person. First, the models will be asked to forecast the deferred Willingness - to - pay, and we will test our models (left panel). Cross-validation will be used to determine the parameters regularization for our Machine Learning models. Data in the training set was divided into seven folds. In contrast, per fold k, the models' mean-squared errors in conjunction with various levels of regularization trained on the other folds were compared. The use of cross-validation will be essential as it will enable the resampling of the data so that it can retrain the data over several iterations (Achar, 2015). For example, the above image splits the data into two, where one part is used for training and learning while the other is used for validating the data and model.

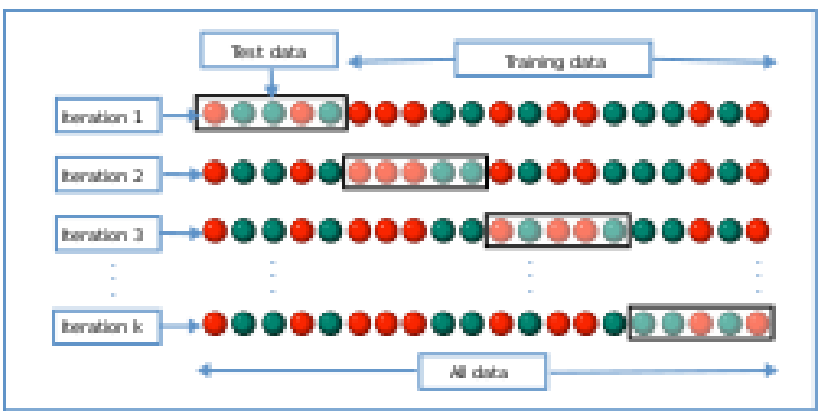

Figure 6: Cross-Validation Iteration

Though machine learning in behavioral science is a form of risk, the dimension associated with risk in deploying machine learning models for economic purposes is counteracted by the use of insurance cover policies. However, the use of machine learning will be a better cause in the deployment of this field (Chimakurthi, 2018).

CONCLUSION

The BDS idea and privacy concerns when governments create AI-based tactics. To achieve this, we thoroughly analyzed the literature and gathered the most significant academic works published. Additionally, we performed semi-structured interviews with professionals giving services to public administrations and governments, and we employed techniques related to data mining to examine the information from these interviews. Finally, based on the findings, we developed a future research plan for the topic under study.

REFERENCES

Achar, S. (2015). Requirement of Cloud Analytics and Distributed Cloud Computing: An Initial Overview. *International Journal of Reciprocal Symmetry and Physical Sciences*, *2*, 12–18. Retrieved from https://upright.pub/index.php/ijrsps/article/view/70

Achar, S. (2016). Software as a Service (SaaS) as Cloud Computing: Security and Risk vs. Technological Complexity. Engineering International, 4(2), 79-88. https://doi.org/10.18034/ei.v4i2.633

Achar, S. (2018). Security of Accounting Data in Cloud Computing: A Conceptual Review. Asian Accounting and Auditing Advancement, 9(1), 60–72. https://4ajournal.com/article/view/70

Barrionuevo, M., Lopresti, M., Miranda, N., & Piccoli, F. (2018). Secure computer network: Strategies and challengers in big data era. *Journal of Computer Science and Technology*, *18*(3). https://doi.org/10.24215/16666038.18.e28

Chimakurthi, V. N. S. S. (2017). Cloud Security - A Semantic Approach in End to End Security Compliance. Engineering International, 5(2), 97-106. https://doi.org/10.18034/ei.v5i2.586

Chimakurthi, V. N. S. S. (2018). Emerging of Virtual Reality (VR) Technology in Education and Training. Asian Journal of Humanity, Art and Literature, 5(2), 157-166. https://doi.org/10.18034/ajhal.v5i2.606

Dhall, S., Bhushan, B., & Gupta, S. (2016). An improved hybrid mechanism for secure data communication. *International Journal of Computer Network and Information Security*, *8*(6), 67.

Emani, S., & Yamin, C. K. (2012). Patient perceptions of a personal health record: A test of the diffusion of innovation model. *Journal of Medical Internet Research*, *14*(6). https://doi.org/10.2196/jmir.2278

Gao, R., Wen, Y., & Zhao, H. (2015). Secure data fusion in wireless multimedia sensor networks via compressed sensing. *Journal of Sensors*. https://doi.org/10.1155/2015/636297

Gao, R., Wen, Y., Zhao, H., & Meng, Y. (2014). Secure data aggregation in wireless multimedia sensor networks based on similarity matching. *International Journal of Distributed Sensor Networks*, https://doi.org/10.1155/2014/494853

Pahl, C., & Xiong, H. (2013, September 1). Migration to PaaS clouds - Migration process and architectural concerns. *2013 IEEE 7th International Symposium on the Maintenance and Evolution of Service-Oriented and Cloud-Based Systems*, IEEE Xplore, 86-91. https://doi.org/10.1109/MESOCA.2013.6632740

Pham, T. D., Tran, D., & Ma, W. (2017). Ownership protection of outsourced biomedical time series data based on optimal watermarking scheme in data mining. *Australasian Journal of Information Systems, 21.* https://doi.org/10.3127/ajis.v21i0.1541

Talukder, A. K., Zimmerman, L., & A, P. H. (2010). Cloud Economics: Principles, Costs, and Benefits. *Computer Communications and Networks*, 343–360. https://doi.org/10.1007/978-1-84996-241-4_20

Yang, Z., Huang, Y., Li, X., & Wang, W. (2018). Efficient secure data provenance scheme in multimedia outsourcing and sharing. *Computers, Materials, & Continua, 56*(1), 1-17. https://doi.org/10.3970/cmc.2018.03697

Zentall, T. R., Galizio, M., Critchfield, T. S. (2002). Categorization, concept learning, and behavior analysis: An introduction. *Journal of the Experimental Analysis of Behavior, 78*(3), 237-248.

Zhuoxuan, J., Yan, Z., Xiaoming, L. (2015). Learning behavior analysis and prediction based on MOOC data. *Journal of computer research and development, 52*(3), p. 614.

--0--

.