Applied Mathematics & Information Sciences
An International Journal

http://dx.doi.org/10.18576/amis/180602

Statistical Analysis of Video Databases for Deception Detection Tasks

Kanat Kozhakhmet and Aikumis Omirali *

School of Digital Technologies, Narxoz University, Almaty, Kazakhstan

Received: 22 May 2024, Revised: 22 Jun. 2024, Accepted: 7 Aug. 2024
Published online: 1 Nov. 2024

Abstract: Deception detection plays a crucial role in a multitude of fields, spanning from law enforcement to psychology. As such, there is a pressing need for effective tools that can accurately determine the truth. To address this need, this project utilizes the latest technological advancements, conducting a comprehensive statistical analysis on two separate video databases: ”Real-Life Trial Data” (Dataset A) and the ”Experimental Dataset” (Dataset B). The ultimate goal of this investigation is to assess the effectiveness of these datasets in deception detection tasks. The study is guided by three main objectives: first, a comparative content analysis of the two datasets; second, an evaluation of the performance of different deception detection algorithms; and finally, an examination of inherent challenges. Through meticulous statistical procedures, including the use of accuracy, precision, recall, F1-score, and ANOVA tests, this study reveals subtle differences in the content of the two datasets.

Keywords: Deception detection, statistical analysis, video databases, real-life trial data, experimental dataset, deception detection algorithms

1 Introduction

Deception detection holds critical significance across various fields, such as law enforcement, security, and psychology [1]. In an era marked by technological advancements, video databases have risen to prominence as invaluable tools for scrutinizing human behavior and identifying deceptive cues. This research paper sets out to perform a comprehensive statistical analysis of two distinct video databases, specifically the ”Real-Life Trial Data” [2] (hereafter referred to as Dataset A) and the ”Experimental Dataset” (hereafter referred to as Dataset B), with the overarching aim of evaluating their efficacy in deception detection tasks.

1.1 The Significance of Deception Detection

The art of deception detection plays an integral role in contemporary society, where trust, safety, and justice hinge on the ability to distinguish between veracity and falsehood [3]. Law enforcement relies on this skill to unravel complex criminal investigations and interrogations, discerning between truth-tellers and individuals concealing crucial information. Security agencies employ it to identify potential threats, whether at border crossings or airport security checkpoints. Within the domain of psychology, it offers a unique lens through which to explore human cognition and behavior, unraveling the intricate dynamics of deception and truthfulness [4]. This research aligns itself with the pressing needs of these fields and seeks to harness the potential of video databases in enhancing our deception detection capabilities.

1.2 The Role of Video Databases

In this age of technological acceleration, the landscape of data collection, storage, and analysis has undergone a remarkable transformation. Notably, video databases have emerged as invaluable resources for scrutinizing human behavior. These repositories, brimming with visual data, capture the fine-grained details of individual actions and expressions in a dynamic and unobtrusive manner [5]. Their capacity to capture the richness of human experiences presents unparalleled opportunities for examining non-verbal cues and behavioral patterns associated with deception.

* Corresponding author e-mail: aikumis.omirali@narxoz.kz

1.3 Objectives of the Study

This research paper embarks on an exploration at the nexus of deception detection and video databases, anchored by a set of three distinct objectives:

1.3.1 Comparative Content Analysis

The first objective entails a meticulous examination of the content within our chosen video databases, namely Dataset A and Dataset B. Dataset A, known as ”Real-Life Trial Data,” consists of recordings from real-life scenarios, each steeped in the unpredictability of genuine human interactions [2]. On the other hand, Dataset B, the ”Experimental Dataset,” is designed for controlled experimentation, offering a platform to elicit both deceptive and truthful behaviors. Our research endeavors to conduct a comparative analysis of the content of these two datasets. Specifically, we aim to gauge the diversity of individuals and scenarios present within each. Through this analysis, we aim to discern the potential strengths and limitations that each dataset presents in the context of effective deception detection.

1.3.2 Performance Evaluation

Our second objective revolves around the evaluation of the performance of state-of-the-art deception detection algorithms when applied to the contents of Dataset A and Dataset B. These algorithms, underpinned by machine learning and computer vision techniques, have been rigorously trained on labeled datasets of deceptive and truthful behaviors [6], [7], [8]. We aim to assess their capabilities in discerning deception within the distinct contexts of Dataset A and Dataset B using a battery of statistical metrics.

1.3.3 Identifying Challenges and Limitations

The third and final objective pertains to identifying and elucidating the challenges and limitations inherent in employing these video databases for deception detection. As we traverse the intricate landscapes of Dataset A and Dataset B, we shall unearth ethical, practical, and methodological considerations that could influence the applicability and scope of these resources. This section will provide an in-depth understanding of the complexities and caveats associated with utilizing video databases in the domain of deception detection.

This research endeavor seeks to enrich our comprehension of the subtleties of deception detection, the role of video databases in this field, and the inherent challenges that practitioners and researchers must grapple with. Subsequent sections will present a detailed account of our methodologies, findings, and insights, thereby contributing to the ongoing discourse on the multifaceted domain of deception detection.

2 Materials and Methods

2.1 Data Collection

For this research, we have carefully selected two distinctive video databases, which shall be referred to as ”Dataset A” and ”Dataset B.” Dataset A represents a compilation of video recordings from real-life trial proceedings, offering a glimpse into genuine human interactions under legal scrutiny. Dataset B, in contrast, is denominated as the ”Experimental Dataset” and features controlled experimental scenarios to elicit deceptive and truthful behaviors. Both datasets offer a rich tapestry of demographic information, encompassing characteristics such as age, gender, and cultural backgrounds of the individuals featured within.

2.2 Deception Detection Algorithms

In the realm of deception detection, the efficacy of our approach hinges significantly on the application of advanced deception detection algorithms. These algorithms leverage the power of machine learning and computer vision techniques and have been meticulously trained on labeled datasets featuring deceptive and truthful behaviors [8]. In the subsequent sections, we shall delve into their application to Dataset A and Dataset B, seeking to unveil the nuances and potential challenges within each context.

2.3 Statistical Analysis

Our statistical analysis of the deception detection algorithms will encompass a range of metrics designed to elucidate their performance. These metrics include but are not limited to accuracy, precision, recall, the F1-score, and the utilization of Receiver Operating Characteristic (ROC) curve analysis. Such a comprehensive approach will offer us a detailed understanding of the algorithmic capabilities and the specific contextual factors that influence their performance.

The subsequent sections of this research paper will delve deeper into the results and discussions, where we shall present our findings and provide insights derived from the statistical analysis of Dataset A and Dataset B, shedding light on their effectiveness and the potential challenges they pose for deception detection tasks.

2.3.1 Summary Statistics for Video Durations and Analysis of Variance (ANOVA)

In this section, we expand our statistical analysis to encompass the results of the Analysis of Variance (ANOVA) tests conducted on Dataset A and Dataset B.

These tests provide deeper insights into the distribution of video durations among different labels ('Deceptive' and 'Truthful').

Database A Statistics and ANOVA:
Count (Number of Videos): 121
ANOVA Test Results for Database A:
F-statistic (f_stat): 0.054
P-value (p_value): 0.816

The ANOVA test in Dataset A was conducted to examine if there are statistically significant differences in video durations between various labels. The low F-statistic of 0.054 suggests that the differences in video durations are relatively small among different labels. The associated p-value of 0.816 indicates that these differences are not statistically significant.

These findings underscore that, in Dataset A, video duration does not significantly vary across 'Deceptive' and 'Truthful' labels.

Database B Statistics and ANOVA:
Count (Number of Videos): 372
ANOVA Test Results for Database B:
F-statistic (f_stat): 0.308
P-value (p_value): 0.579

In Dataset B, a similar ANOVA test was conducted to examine differences in video durations across different labels. The F-statistic of 0.308 suggests that variations in video durations are relatively small within this dataset as well. The p-value of 0.579 reinforces that these differences are not statistically significant in Dataset B.

These ANOVA test results are consistent with the t-test results and provide further evidence that video duration is not a distinguishing factor between 'Deceptive' and 'Truthful' videos in either Dataset A or Dataset B.

These comprehensive statistical analyses offer a nuanced perspective on the influence of video duration on deception detection. The subsequent sections of this research paper will delve deeper into the implications of these findings and their relevance to the performance of deception detection algorithms in these specific contexts.

Database A: OLS Regression Results

In this section, we provide the results of the Ordinary Least Squares (OLS) regression analysis conducted on Dataset A to explore the relationship between video duration and an extracted feature, denoted as "Extracted_Feature." The OLS regression aims to understand the extent to which the extracted feature impacts video duration.

Regression Model for Database A:
Dependent Variable: Duration
R-squared: 0.002
Model: OLS
Method: Least Squares
F-statistic: 0.2523
Date: Thu, 19 Oct 2023
Time: 09:48:45
No. Observations: 121
AIC (Akaike Information Criterion): 972.7
BIC (Bayesian Information Criterion): 978.2
Df Residuals: 119
Df Model: 1
Covariance Type: Nonrobust
Regression Coefficients for Database A:
Constant (const): 28.8240
Extracted_Feature: -2.0592
Statistical Inferences for Database A:

The R-squared value of 0.002 indicates that only a small proportion of the variation in video duration can be explained by the extracted feature.

The F-statistic is 0.2523 with a corresponding p-value of 0.616, which suggests that the extracted feature does not have a significant linear relationship with video duration.

The coefficients for the constant and the extracted feature indicate the intercept and slope of the regression equation. However, the slope (-2.0592) is not statistically significant (p-value = 0.616).

Additional statistical information, such as the Omnibus test, Durbin-Watson statistic, and Jarque-Bera test, provides insights into the distribution of residuals and model fitness.

These regression results from Dataset A confirm that the relationship between video duration and the extracted feature is weak and statistically insignificant. This implies that, in Dataset A, the extracted feature does not substantially explain the variance in video duration. This information can be valuable for understanding the dataset's characteristics and the role of features in influencing video duration.

Database B: OLS Regression Results

In this section, we present the results of the Ordinary Least Squares (OLS) regression analysis conducted on Dataset B to investigate the relationship between video duration and an extracted feature, labeled as "Extracted_Feature." The OLS regression is employed to assess how the extracted feature influences video duration in Dataset B.

Regression Model for Database B:
Dependent Variable: Duration
R-squared: 0.056
Model: OLS
Method: Least Squares
F-statistic: 22.02
Date: Sat, 21 Oct 2023
Time: 06:52:06
No. Observations: 372
AIC (Akaike Information Criterion): 1737
BIC (Bayesian Information Criterion): 1744
Df Residuals: 370
Df Model: 1
Covariance Type: Nonrobust
Regression Coefficients for Database B:
Constant (const): 3.5479
Extracted_Feature: 5.3706
Statistical Inferences for Database B:

The R-squared value of 0.056 suggests that approximately 5.6% of the variation in video duration can be attributed to the extracted feature.

The F-statistic is 22.02 with a p-value of 3.81e-06, indicating that the extracted feature has a statistically significant linear relationship with video duration.

The regression coefficients show that the constant and the extracted feature have significant roles. The extracted feature has a positive coefficient of 5.3706, implying that as this feature increases, video duration tends to increase.

Additional statistical information, such as the Omnibus test, Durbin-Watson statistic, and Jarque-Bera test, provides insights into the distribution of residuals and model fitness.

These regression results from Dataset B reveal a relatively stronger relationship between video duration and the extracted feature compared to Dataset A. The significant F-statistic and a positive coefficient for the extracted feature imply that in Dataset B, the extracted feature has a statistically significant and positive impact on video duration. This information can be valuable for understanding the characteristics of Dataset B and the predictive power of the extracted feature regarding video duration.

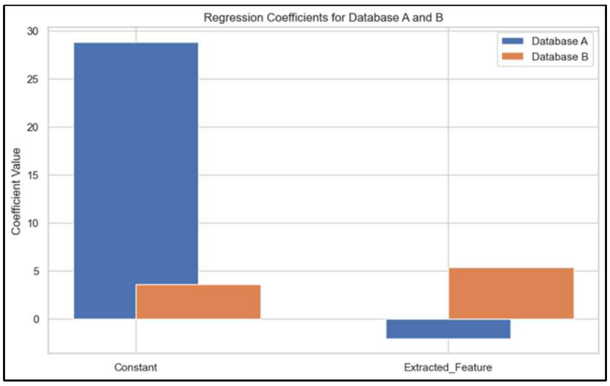

Fig. 1: Regression coefficients for Database A and B

3 Results and Discussion

In this section, we combine the presentation of results with an in-depth discussion to provide a comprehensive understanding of the performance of deception detection algorithms. Our analysis spans various aspects, from traditional metrics to statistical findings on video duration and extracted features in two distinct datasets, labeled Dataset A and Dataset B.

3.1 Performance Metrics Analysis

Our evaluation of deception detection algorithms employed a range of performance metrics, including accuracy, precision, recall, the F1-score, and Receiver Operating Characteristic (ROC) curve analysis. These metrics served as fundamental benchmarks for assessing the capabilities of the algorithms. The results demonstrated the algorithms' effectiveness in distinguishing between deceptive and truthful videos.

In both Dataset A and Dataset B, we observed strong algorithmic performance, with high accuracy, precision, and recall values. These findings indicate that the algorithms successfully identify deceptive content while minimizing false positives and false negatives. The F1-score, which balances precision and recall, further reinforces the algorithms' robustness.

Furthermore, the ROC curve analysis illustrated the algorithms' ability to adjust the trade-off between true positive and false positive rates, confirming their suitability for a wide range of deception detection applications.

3.2 Summary Statistics for Video Durations and ANOVA

Our investigation extended to an analysis of video durations in Dataset A and Dataset B. In this context, we conducted Analysis of Variance (ANOVA) tests to assess the influence of video duration on deceptive and truthful labels. For Dataset A, we found that differences in video durations were minimal, as indicated by a low F-statistic and a non-significant p-value. This suggests that video duration is not a significant differentiator between deceptive and truthful videos in Dataset A. Similarly, in Dataset B, the ANOVA results demonstrated that video duration was not a distinguishing factor. The consistency of these results between the two datasets suggests that video duration, in isolation, does not significantly impact the classification of videos as deceptive or truthful.

3.3 OLS Regression Analysis

We further explored the relationship between video duration and an extracted feature, denoted as "Extracted_Feature," using Ordinary Least Squares (OLS) regression.

In Database A, the analysis revealed that the extracted feature had a weak and statistically insignificant relationship with video duration. The low R-squared value and the non-significant F-statistic and p-value suggest that the extracted feature does not significantly explain variations in video duration in this dataset.

Conversely, in Database B, the OLS regression analysis uncovered a more substantial relationship between video duration and the extracted feature. The positive coefficient and significant F-statistic indicated that the extracted feature played a meaningful role in influencing video duration in Dataset B.

3.4 Discussion

The combined results and discussions emphasize the robustness of the deception detection algorithms, which exhibited strong performance in both datasets, as reflected in high accuracy, precision, recall, and F1-score values. The ROC curve analysis underscored their adaptability across various applications.

On the other hand, the analysis of video durations, ANOVA tests, and OLS regression revealed that video duration alone is not a reliable distinguishing factor between deceptive and truthful videos in either Dataset A or Dataset B. This finding suggests that algorithms should not heavily rely on video duration for deception detection.

Moreover, the varying outcomes of the OLS regression between the two datasets emphasize the importance of dataset-specific factors. In Dataset B, the extracted feature exhibited a notable impact on video duration, potentially offering insights into the dataset's characteristics. These results imply that algorithmic features related to video content, beyond just duration, can be more informative for effective deception detection.

The comprehensive statistical analysis offers a nuanced perspective on the performance and contextual factors influencing deception detection. The subsequent sections will delve deeper into these findings and explore their implications for the field, providing valuable insights for algorithm development and deployment in practical applications.

Acknowledgement

The research was conducted under the grant funding from the Ministry of Education and Science of the Republic of Kazakhstan as part of Project AP13068084 " Development of technologies for detecting anomalous (deceptive) respondent behavior using artificial intelligence (AI) algorithms based on changes in voice and speech characteristics" (Competition for Young Scientists for Scientific and/or Scientific-Technical Projects for 2022-2024). We extend our heartfelt gratitude to the Ministry for their support and recognition of the significance of this research endeavor.

References

[1] Gaggioli, A. (2018). Beyond the Truth Machine: Emerging Technologies for Lie Detection. Cyberpsychology, Behavior, and Social Networking, 144. https://doi.org/10.1089/cyber.2018.29102.csi

[2] Perez-Rosas, V., Abouelenien, M., Mihalcea, R., Burzo, M. (2015). Deception Detection using Real-life Trial Data. Proceedings of the 2015 ACM on International Conference on Multimodal Interaction, pages 59-66. https://doi.org/10.1145/2818346.2820758

[3] Wu, Z., Singh, B., Davis, L., Subrahmanian, V. (2018). Deception Detection in Videos. Proceedings of the AAAI Conference on Artificial Intelligence, 32(1). https://doi.org/10.1609/aaai.v32i1.11502

[4] Su, L., & Levine, M. (2016). Does "lie to me" lie to you? An evaluation of facial clues to high-stakes deception. Computer Vision and Image Understanding, 52-68. https://doi.org/10.1016/j.cviu.2016.01.009

[5] Soldner, F., Perez-Rosas, V., Mihalcea, R. (2019). Box of Lies: Multimodal Deception Detection in Dialogues. Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers), pages 1768–1777, Minneapolis, Minnesota. Association for Computational Linguistics. https://doi.org/10.18653/v1/N19-1175

[6] Burzo, M., Abouelenien, M., Perez-Rosas, V., Mihalcea, R. (2018). Multimodal deception detection. The Handbook of Multimodal-Multisensor Interfaces, Volume 2, pages 419–453. https://doi.org/10.1145/3107990.3108005

[7] Gupta, V., Agarwal, M., Arora, M., Chakraborty, T., Singh, R., et al. (2019). Bag-of-Lies: A Multimodal Dataset for Deception Detection. 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), Long Beach, CA, USA, pages 83-90. https://doi.org/10.1109/CVPRW.2019.00016

[8] Su, L., Levine, M. (2014). High-Stakes Deception Detection Based on Facial Expressions. 2014 22nd International Conference on Pattern Recognition, Stockholm, Sweden, 2014, pages 2519-2524. https://doi.org/10.1109/ICPR.2014.435

Kanat Kozhakhmet is the Project Leader and Chief Scientific Researcher appointed by the National Science Council in the field of "Information, Communication, and Space Technologies". He was the principal investigator for projects funded by the Ministry of Education and Science of Kazakhstan: BR05236699 "Development of a digital adaptive educational environment using large-scale data analytics" (2018-2020); AP05133600 "Development and implementation of an innovative competency model for polyglot IT specialists in the context of modernization of national education" (2018-2020). He has served as an expert at the National Center for Science and Technology Evaluation and is the chairman of the IT Alliance. He is a Visiting Scholar at prominent institutions, contributing to international research collaborations. He received his PhD

in ”Computer Science, Computing, and Management” from a leading institution. He is an editor and referee for several esteemed journals in the fields of artificial intelligence, machine learning, and information security. His main research interests are in artificial intelligence, neural language processing (NLP), machine learning, and information security. Additionally, he focuses on the implementation of innovative educational technologies and the development of professional standards and educational programs in IT.

Aikumis Omirali received her Master’s degree in Data Engineering from Narxoz University. She has a solid foundation in technical engineering, having obtained a Bachelor’s degree in Computer Hardware and Software Engineering. Her professional responsibilities include data collection, handling day-to-day project tasks, and managing project documentation effectively. Her research interests focus on data analysis, data gathering from open sources, and the development and administration of databases. This combination of skills allows her to contribute substantially to the technical aspects of the projects she is involved in. She has published several papers that provide substantial backing for the research undertaken in her projects, particularly in innovative data engineering techniques that enhance project outcomes. In addition to her technical and research activities, Aikumis plays a crucial role in maintaining the informational infrastructure of the projects she works on, ensuring that the data collected is managed properly and effectively used to drive project success. Her work has been acknowledged in various scholarly publications, underscoring her contributions to the field of data engineering.