Investigating Computer Vision-Based Approaches in Lie Detection

Ara Mambreyan

JDF 2020/2021 Project Summary

Background

Lying is common in our daily lives. Often, lies are innocuous such as the existence of Santa Claus. However, some lies can ruin lives and pose an existential threat to society.

Lying in court affects the justice system, allowing guilty defendants to go free and innocents imprisoned. Furthermore, there has been a global rise in terrorism in recent years. Fake news is becoming increasingly prevalent in our daily lives.

An accurate lie detector can revolutionize the justice system, combat terrorism and minimize the spread of fake news. With the recent advances in Artificial Intelligence (AI), new machine learning methods for lie detection have been researched which use various features such as video (face, body), audio (voice) and speech (language) to detect lying.

Recent techniques are based on facial expressions known as "micro-expressions" (Figure 1). Micro-expressions reflect emotions that subjects might want to hide which can be used for predicting deception.

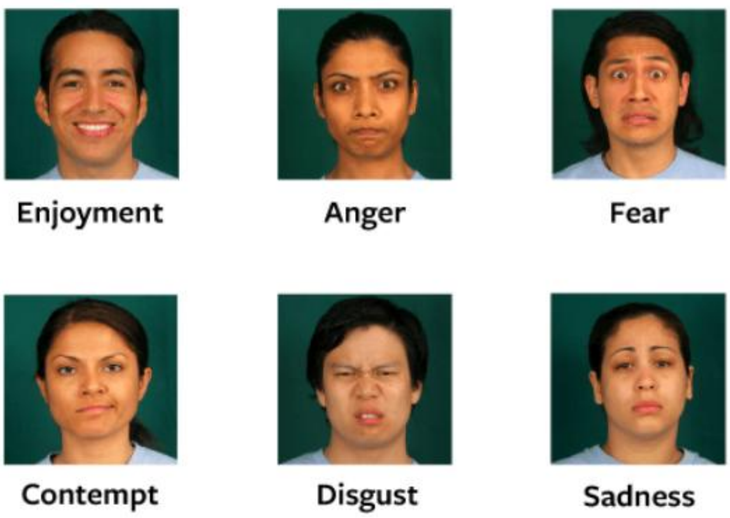

Figure 1: Illustration of several micro-expressions.

Using the concept of micro-expressions, techniques in research achieved extremely high results (99% accuracies). Unsurprisingly, all of this gave rise to the media hype of AI LIe Detectors e.g. in outlets such as The Guardian and WIRED Magazine.

However, in the rush to create the perfect AI Lie Detector, ethics was left behind. In our project, we evaluate common pitfalls which could have led to existing techniques overestimating their results. In particular, we investigate gender bias.

Dataset

The dataset used in the papers we are critiquing and our experiments consists of video, audio, speech and labels of "Truth" or "Lie". The videos are from a real-life trial dataset, acquired from public courtroom hearings. The dataset is also manually annotated with micro-expressions for each video.

The dataset consists of 121 videos from real-life courtroom hearings with 61 lies and 60 truths from 56 subjects. The videos are shot in an unconstrained setting with significant variations in pose, illumination and size (Figure 2).

Figure 2: Screenshots from the Real-life Trial dataset videos.

To investigate gender bias, we examine the percentage of lies for two genders, female and male. Table 1 shows the data point distribution from females and males, along with the percentage of lies. Clearly, in this particular dataset, females lie more often than males. This means that Machine Learning algorithms could exploit this defect and achieve overly optimistic results.

Gender	# of Points	Lie %
Female	76	64.5%
Male	45	26.7%

Table 1: Real-life Trial dataset statistics for gender. The second column is the number of data points and the last column is the percentages of lies.

Experiments

To test the effect of gender bias, we conduct two experiments. Firstly, we train a Machine Learning algorithm on gender. Then, when the algorithm needs to make a prediction of truthfulness, it predicts gender. If the predicted gender is "Female", it assigns the label "Lie" and vice-versa. Figure 3 shows the diagram for this.

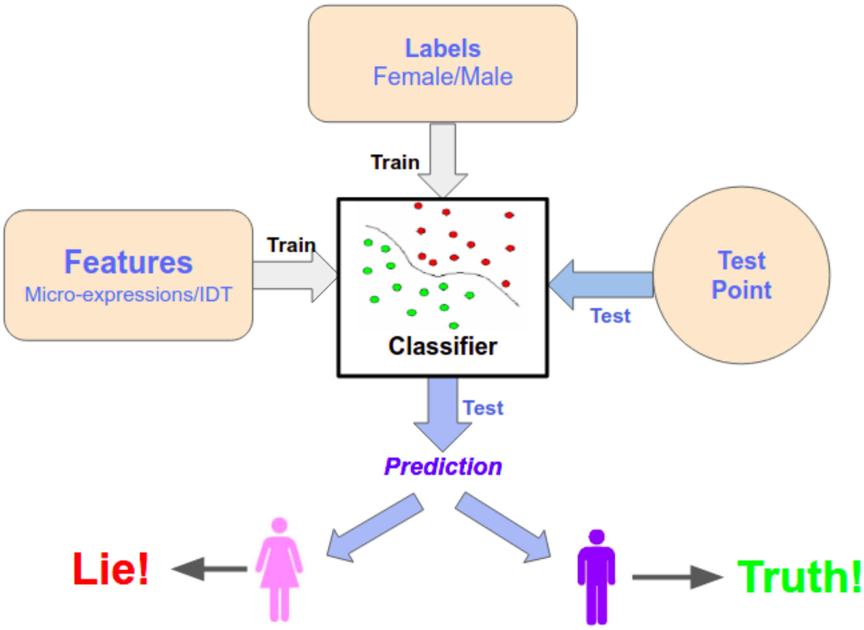

Figure 3: Illustration of the experiment performed to show gender bias. The algorithm is first trained with gender labels, "Female" and "Male". For a new test data point, the classifier predicts the gender. If the prediction is "Female", the classifier assigns "Lie" and if the prediction is "Male" the classifier assigns "Truth".

ACC	LR	KNN	L-SVM	K-SVM	RF	MLP	AB
L	72	72	76	73	78	74	**78**
G	63	60	64	63	63	**65**	62

Table 2: The accuracy scores achieved with different Machine Learning algorithms.

The ad-hoc method is compared with an algorithm actually trained on truth/lie labels. If the ad-hoc method achieves statistically significant results or, better, comparable results to an algorithm trained on truth/lie labels, we could conclude that gender bias is significant.

We perform the experiment using manually annotated micro-expressions as features. Table 2 shows the obtained results with lies (row 1) and gender (row 2).

It can be clearly seen that the results achieved with gender are statistically significant and not too far from the algorithms trained with truth/lie labels. This means that the gender bias effect is significant.

Conclusion

Besides manually annotated micro-expressions, we also performed experiments using other features obtained from video such as IDT which further showed that the gender bias effect is large. In addition, experiments were performed on an unbiased dataset and results were no better than a coin toss. All of this suggests that there is no evidence that AI Lie Detection actually works.

We believe our findings will save time for researchers who are working with the assumption that current state-of-the-art papers in lie detection are valid and achieve almost perfect accuracies. Finally, we hope that our findings will recalibrate expectations of AI Lie Detection technology in both industry and research communities alike.