A generative framsework for modeling human behavior with an application to analysis of deceptive behavior

Nikolas Marcou*, Stefanos Zafeiriou* and Maja Pantic *#
* Dept. of Computing, Imperial College London 180 Queen’s Gate, London SW7 2AZ, U.K. {s.zafeiriou,gt204,m.pantic}@imperial.ac.uk
# EEMCS University of Twente Drienerlolaan 5 7522 NB Enschede, The Netherlands

Abstract

Deception is a message knowingly and intentionally transmitted with the intent to foster false beliefs or conclusions. This definition encompasses strategies to mislead, such as equivocation, ambiguity, evasiveness and outright falsification. There are differences between deception and lying, but they have been ignored by the majority of research. However, wherever two people communicate, deception is a reality. It is present in our everyday social and professional lives and its detection can be beneficial, not only to us individually but to our society as a whole. For example, accurate deception detection can aid law enforcement officers in solving a crime. It can also help border control agents to detect potentially dangerous individuals during routine screening interviews. In this paper we propose a generative approach for learning the underlying dynamic structure of behavioral patterns and we apply it for recognizing high level tasks such as the difficulty of a question asked in a show and analysis of deceptive behavior.

1. Introduction

In [1] Ekman and Friesen published the theory about deception. They suggested that cues fell under two categories, leakage cues and deception cues. Leakage describes the movements that reveal how most people feel even when they attempt to conceal that information. By analyzing the experimental videotapes of people lying and telling truth, they found instances in which the muscles activity was not inhibited. These actions were then called "reliable" facial muscles. They queried whether face or body is more reliable for deception detection. Darwin thought people could comment on their body movements so that they should be easy to conceal, unlike facial expressionss. They summarized it as "face - body leakage" hypothesis. Although body movements (hands and feet) would be relatively easy to inhibit, most people do not bother to censor their body movements due to the lack of feedback from others. On the other hand, people usually do not fine-tune their body actions when they lie. So they suggested adding bodily movements as one of the reliable cues in combination with vocal behaviors.

Deception, although having been linked to a number of observable behaviors and physiological reactions, cannot be directly measured. However, the arousal, cognitive load and self monitoring that may accompany deception can. Stress has long been associated with deceptive behavior during interrogation. It emanates from the brain center and manifests itself in the peripheral senses through a variety of physiological signatures [2][3]. It is through these cues that lie-catchers ferret out deception.

Trying to find non verbal indicators of deception has been of interest to the scientists for the last 30 years. Currently, the most successful and widespread system is the polygraph which monitors uncontrolled changes in heart rate and electro-dermal response, as a result of the subject's arousal to deceit. Contact sensors are placed on the subjects during interrogations. The resulting signals are then heuristically analyzed by experts [4]. Such a complex process comes with various disadvantages.

(a) The comfort of subjects is compromised, which is very important in psycho-physiology, as it is believed to contaminate the experiment [5].

(b) The person examined must be fairly cooperative and in close proximity to the device.

(c) The analysis of the polygraph results can be quite slow and labor intensive as it is manually performed by several experts who later cross exam their findings. The processing of a typical 10 minutes interrogation session may take several hours.

(d) In addition, due to the heuristic nature of the analysis, the outcome cannot be embraced with high confidence and always has a suggestive nature [6].

(e) Even though large companies and agencies pass some of their employees in certain positions through a polygraph examination, the latter is by no means a procedure that could be used on all people holding sensitive information and would certainly not be used casually by law enforcement on their daily duties.

(f) Invasive procedures have by nature a limited use in our society where comfort levels of people are to be respected.

Having conducted a set of laboratory tests, Vrij [7] suggested that the polygraph is about 82% accurate at identifying deceivers. The National Academy of Sciences, however, concluded that such experimental numbers are often overestimates of actual results, especially in personnel screening [6]. Not all deception detection methods are invasive. Computerized Voice Stress Analysis (CVSA), for example, is a technique that analyzes voice pitch changes as a measure of arousal. The technique has shown to be roughly equivalent in accuracy to the polygraph, but as with the polygraph [8], this method will not be useful in situations where deception is not accompanied by physiological arousal [9]. Therefore there exists an increasing demand for real-time, non invasive, automated methods for deception detection. Methods that would be based on repeatable computer algorithms and that would require the minimum cooperation from the person being interviewed are essential.

Previous works employed several cues like hand movement, thermal imaging and eye tracking (discussed in detail later on). The results were encouraging but limited as determining possible human intention only from observable cues is not only a difficult endeavor but can also be impossible. According to Ekman, "there is no clue to deceit that is reliable to all human beings" [10]. Bond and Fahey [11] also proved that nervous truth tellers might show the same nervous behaviors as liars. The fear of not being believed (truth tellers) and the fear of getting caught (liars) will produce the same behaviors, namely signs of nervousness.

Although seemingly de-motivational the previous statements make the work we set out to do much clearer. In our quest for truth we will focus on physical manifestations that are hard to pretend and attempt to detect the presence of stress. The variety of intentions is limitless and an individual may possess multiple intentions at any point in time, so an automated detector that relies on stress cues can only provide strong suggestion (and not proof) of deceit.

Some of the recent advances in automated verity/deceit decision-making include a computer-based linguistic analysis for deception detection [12] which used decision trees to show that deceivers often displayed higher quantity of information and expressiveness and used less vocabulary and grammar [13]. Another pattern recognition based decision classifier was based on thermal imaging analysis [14] in which body heat changes were measured via thermal signals while subjects were being interviewed. Lastly, an automated deceit detection technique was developed by Nwogu et al. [15] where the authors extracted several features such as blink rate, gaze duration and manually clustered the features. They subsequently tagged each cluster as truth or deceit and reported a deceit detection accuracy of only 64.28%.

In [13] the authors presented a prototype for an automated deception detection system. They investigated the role of dynamic eye-based features such as eye closure/blinking and lateral movements of the iris in detecting deceit. The features were recorded both when the test subjects were having non-threatening conversations as well as when they were being interrogated about a crime they might have committed. The rates of the behavioral changes were blindly clustered into two groups. Examining the clusters and their characteristics, they observed that the dynamic features selected for deception detection show promising results with an overall deceptive/non deceptive prediction rate of 71.48% from a study consisting of 28 subjects.

In [16] the authors proposed a novel approach for deriving indicators of deception from video-taped interactions. Their method focuses on deriving cues from the head and hands since these areas are a proven source of reliable indicators of deception [17]. General metrics were extracted from the video using a method called "blob analysis". This method uses color analysis, eigenspace-based shape segmentation and Kalman filters to track head and hand positions throughout the video segment. Their final model included average and variance of the head position and angle, the average and variance of the positions of the head and hands, the average distance between the hands, the average and variance of the distance between the center of the triangle and the hands and the head, the average triangle area, the variance of the center position of the triangle and the average number of frames the hands are located in each quadrant. Using discriminant analysis they classified the deceptive and truthful participants with an accuracy rate of 89.5%. However, when one participant was withheld from the analysis (leave one out, cross validation) and was used for testing, the accuracy decreased to 60.5%.

Previous works have demonstrated the correlation of increased blood perfusion in the orbital muscles and stress levels for human beings. It has also been suggested that this periorbital perfusion can be quantified through the processing of thermal video. The idea has been based on the fact that skin temperature is heavily modulated by superficial blood flow. In [14] the authors proposed a new methodology to compute the mean periorbital temperature signal. This methodology featured a tandemn condensation tracker to register the periorbital area in the context of a moving face. It operated on the raw temperature signal and tried to improve the information content by suppressing the noise level instead of amplifying the signal as a whole. Finally a pattern recognition method classified stressful (deceptive) from non-stressful (non-deceptive) subjects based on a comparative measure between the entire interrogation signal and a critical subsection of it. The successful classification rate was 87.2% for 39 subjects [14].

Although there have been many research works in systemically detecting deceit in the behavioral psychology community, automation of the deceit detection processes is still in its infant stage [13]. In this paper we propose a generative approach for modeling complex dynamic phenomena such as human facial behavior. In the majority of works that apply such a modeling, features are directly used in a dynamic inference model (such as Hidden Markov Model (HMM) [18]). In this paper we take a different track, instead of using raw features we apply an unsupervised procedure for learning a basic alphabet in order to represent the observable symbols of the HMM. We apply this modeling for learning the dynamic structure of high level tasks such as the difficulty of a question asked in a show and the analysis of deceptive behavior.

2. Data Preparation

2.1. The moment of truth

The "Moment of Truth" is a television game show in which contestants answer a series of 21 increasingly personal and embarrassing questions to receive cash prizes. The show is hosted by Mark L. Walberg and aired on the Fox network. The show premiered on January 23, 2008 and ended on August 8, 2009[19][20].

Prior to the show, each contestant is administered a polygraph exam and asked more than 50 questions - many of which are then asked again in front of the studio audience during the actual taping of the program. Without knowing the results of the polygraph, he or she is asked 21 of those same questions again, each of which is progressively of more personal nature. If the contestant answers honestly, according to the polygraph results, he or she moves on to the next question; however, should a contestant lie in his or her answer (as determined by the polygraph) or simply refuse to answer a question, the game ends. If he/she gives a false answer before the $10,000 level of questions, he/she leaves the show with no prizes [20]. For each tier of questions answered correctly, the contestant wins the corresponding amount of money. A contestant can stop at any time before any question is asked and collect his/her earnings, but once they hear a question, they have to answer it or lose the game. Answering all 21 questions truthfully, as determined by the polygraph results, leads in winning the jackpot of $500,000 [20]. Sometimes, a "surprise guest" (such as an ex-partner or a good friend) can come on stage and ask a particularly difficult question. Friends, colleagues, and family of the contestant who are gathered near the player have access to a button which can be used to switch out a question per game if they feel that the nature of the question is too personal, an option which is introduced to them after the third question [20].

The "Moment of Truth" was selected as a data source for many reasons.

- One finding of deception studies is that the person examined may not feel very much involved in the task, and therefore he/she may not be very likely to produce any nonverbal cues to deception [23]. The show includes a large variety of questions (from amusingly embarrassing during the first questions to deeply personal later on). This is very important as we require varying responses to encompass the range of emotions a person feels when he/she is under stress.
- The show's clean questions format makes it easier to extract data. Questions are asked in a yes or no manner. This makes it very easy to classify given samples. The contestant is asked the question, given time to think about it, responds and then waits for the verdict. While this is taking place we are given a clear view of the contestant's face, making it easy to capture the selected features.
- Last but not least, we are provided with the results from the polygraph examination which is very useful as a benchmark and for validating our results. The data collected were from the first season of the game aired in 2008.

- Last is the lack of publicly available datasets for the task of the analysis of deceptive behaviour.

The raw data source was the whole episodes of the game show in MPEG4 encoded videos. Before we began extracting features we had to pre-process the videos. It was very important to convert all videos to the same resolution and frame rate in order to ensure that the extracted features were on the same scale. Next, we continued by cutting the videos into episodes. Each episode included all the questions of a single player. This was done as each player's questions may span over one or more episodes. Once we had all players’ questions in a single video we cut each question into types. These types were "hear question", "wait", "response", "wait verdict" and "verdict response". The "hear question" type contained the first few seconds where the host asked the player the question. The "wait" type contained the frames where the player was given time to think about it. The "response" type contained the frames where the player answered the questions. In this type we considered only the actual answer and not any talk. The "wait verdict" type contained the frames were the player waited for the verdict to the answer. The final type "verdict response" contained the frames after the player has heard the verdict to his/her answer. We batched all responses of the same type into a large group. It is important that each type contained only clear footage of the player. An Active Shape Tracker (AST) [21][22] was used for tracking a set of facial landmarks in the episodes. The 3d mask mesh had to be manually placed on each player's face (Figure 1 shows an example of such an initialization). The 3D mesh is then tracked in the video sequence (an application of the tracker is shown in Figure 2).

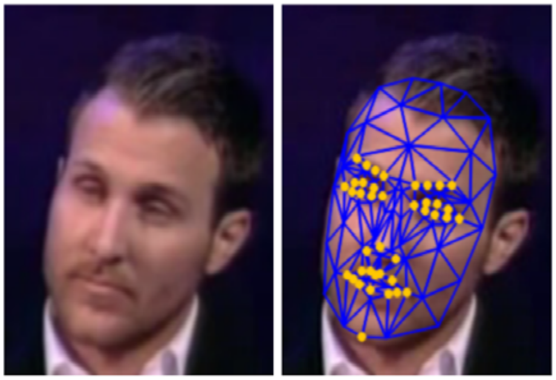

Figure 1: A player's face before and after it is matched by the 3d face mask mesh.

Figure 2: Tracker image output

We began with the 16 originally aired videos and rescaled them so that they all had the same resolution. We then edited them into 918 video sections and dumped those video sections into approximately 100.000 image frames. Subsequently, we used the AST to track the facial points through the images. Finally we projected those points to the appropriate space, rescaled them and extracted the rate of change on the three face axis. The displacements of the facial points were our final features that were used to train our model.

3. Methodology

Hidden Markov models (HMMs) are of great use in problems that have an inherent temporality that is, that they consist of a process that unfolds in time (we have states at time t that are in influenced directly by a state at t-1), such as speech recognition and gesture recognition or in our case facial expressions. Hidden Markov models have a number of parameters whose values are set so as to best explain training patterns for the known category. Later, a test pattern is classified by the model that has the highest posterior probability (i.e. that best "explains" the test pattern).

Our raw feature space is a continuous space of values for our 40 3-dimensional points. However we know that these 40 points don't move independently but have some internal structure that drives them. So we will use k-means clustering to find those clusters of movement and then assign the 120 dimensional data points to the nearest cluster. These clusters will represent the symbols that will be the observables of our Hidden Markov Model. For each of the question's types we will first split it into train and test data. Then we will use k-means to cluster data and train a Hidden Markov Model for deceptive and truthful responses using the Forward-Backward algorithm. We will then attempt to classify the test responses by assigning the sequences to the model that best describes it (i.e. has the highest posterior probability) using the Decoding algorithm.

We use k-means to extract the symbols to be used in Hidden Markov Models. When working on high dimensional data, we are limited in our effort to find an accurate probability distribution by the curse of dimensionality which states that the number of samples must grow exponentially with the dimensionality D of the space. In order to avoid this problem we find the clusters of the most common motions and assign our samples to a set of discrete symbols representing them.

Determining the number of clusters in a dataset, is a frequent problem in data clustering. One method of choosing the number of clusters is to use the Elbow rule which is a heuristic that suggests choosing a number of clusters so that the addition of another cluster doesn't provide a much better modeling of the data. More precisely, if you graph the percentage of variance explained by the clusters against the number of clusters, the first clusters will add much information (explain a lot of variance), but at some point the marginal gain will drop, giving an angle in the graph. Therefore the cluster number is chosen at this marginal point, hence the "elbow criterion". This "elbow" cannot always be unambiguously identified [24]. The percentage of variance is the ratio of the between-group variance to the total variance. We already mentioned that k-means clustering is a heuristic algorithm and thus there is no guarantee for optimality. In order to get good results it is essential to repeat the algorithm with different initial conditions. The number of repetitions is the second parameter of our experiments. Usual repetitions number will be between 10 and 100. The selected distance criterion is the Euclidean distance.

The number of hidden states in the HMM is the main parameter of the model. Each state is a cluster of motions in the 120-dimensional space. Unfortunately the number of states in our problem is not known. In order to determine that number for each response we shall use a simple heuristic. We will iteratively increase the states and stop when the half of the responses show decrease in their log-likelihood. Hidden Markov Models are plagued by the same problem of all learning techniques which is overfitting. Overfitting is controlled by the number of states and clusters. It is tempting to keep increasing the number of states and observe the probability of the sequences rising. But that would be a wrong tactic since it would limit the model's predictive power to the training dataset, especially for the deceit prediction class since we have a limited number of deceitful responses.

In order to evaluate our model we use a technique called cross validation. The principle is that we will split out dataset into a train and a test dataset. We will use the train set to find the model's parameters and the test set to evaluate it. In our case we will use an n-fold cross validation. That means that we will split the dataset into n parts and we will calculate the model n times while withholding a different part each time. That part we will later use for evaluating.

4. Experimental results

We applied the proposed generative framework for modeling the structure of complex dynamic phenomena. The first phenomenon we tried to model was to identify the question a particular response came from (difficulty of the question) given that we provide the classifier with the type of the response. That means we had to train an HMM for each type for each question. Since the maximum question reached in our dataset is 18 the combined HMMs create 5 classifiers (one for each question type) with 14-class output each (question 1 to 14 which corresponds to difficulty 1 to 14). We excluded questions 15, 16, 17 which had only one player sample. The F1 measure was about quite satisfactory and about 69.79%. That means that even though the movements came from different players they all performed facial movements that clustered particularly close.

4.1 Analysis of Deception

The second problem we applied the proposed framework was to classify the responses as deceitful or truthful. The parameters in use were the number of repetitions in the k-means clustering and the accuracy which controls the number of symbols to be used. We learned these parameters using the cross validation framework we implemented. Below follow the confusion matrices for each type:

Response	Truthful	Deceitful
Truthful	84.5%	15.5%
Deceitful	100%	0

Confusion matrix for type ‘wait verdict’.

Response	Truthful	Deceitful
Truthful	80%	20%
Deceitful	100%	0

Confusion matrix for type ‘wait’.

Response	Truthful	Deceitful
Truthful	95%	5%
Deceitful	50%	50%

Confusion matrix for type ‘hear’.

Response	Truthful	Deceitful
Truthful	98%	2%
Deceitful	100%	0%

Confusion matrix for type ‘verdict response’.

Response	Truthful	Deceitful
Truthful	98%	2%
Deceitful	100%	0%

Confusion matrix for type ‘response’.

Even though, the results may not seem very promising there is at least one type for which the recognition rate of deceitful behavior was 50% percent. In the type hear the facial behavior of the contestant immediately after the question is been asked displayed. Of course, deceptive cues may not be evident in all types or these cues maybe more apparent in some types. On the other hand truthful sequences have a very high classification rate. That is mainly due to the large number of responses we had.

5. Discussion- Conclusions

Behavioral psychologists generally believe that attempts to deceit cause physiological reactions such as high blood pressure, increased heart and respiration rate. The physiological reaction is the consequence of arousal that is associated with high-stakes deception [23]. In [25], the authors suggest that the behavioral cues to deceit differ in low- and high-stake situations, i.e. the nervous behaviors manifested or leaked in people telling lies when the stakes are high, are different from when they are low, and high-stake cues are more easily detected. A high-stake lie is one told when the person lying stands to get a notable gain, or faces a notable loss by telling the truth [13]. In our case we examine cases where the players being questioned are between a high monetary reward and personal humiliation for telling the truth and saving their personal image for lying. Our results would therefore not be applicable to all other situations since the personal and economic stakes are relatively high.

The small number of deception samples means that we have not explored the full spectrum of deceptive facial movements. Such a drawback can be overcome only by training our model with a larger sample with more deception samples. It could be interesting to use the same game show from different countries. Given enough samples the results could provide us with some cross-culture deceptive patterns. Finally the majority of the questions are what is called loaded questions, which means that they create pre-assumptions and embarrass the players even prior to answering. Such a fact may be limiting to the predictive power of our model.

6. Acknowledgement

This work was supported by the European Research Council under the ERC Starting Grant agreement no. ERC-2007-StG-203143 (MAHNOB).

8. References

[1] Ekman, P. and W.V. Friesen: The repertoire of nonverbal behavior: Categories, origins, usage, and coding. Semiotica, 1(1):49-98, 1969.

[2] Sokolov, E.N. and J.T. Cacioppo: Orienting and defense rexes: Vector coding the cardiac response. 1997.

[3] Drummond, P.D. and J.W. Lance: Facial ???ushing and sweating mediated by the sympathetic nervous system. Brain, 110(3):793, 1987.

[4] Kleinmuntz, B. and J.J. Szucko: Lie detection in ancient and modern times: A call for contemporary scientific study. American Psychologist, 39(7):766, 1984.

[5] Yankee, WJ: An investigation of sphygmomanometer discomfort thresholds in polygraph examinations. Police, 9(6):12-18, 1965

[6] Stern, PC: The polygraph and lie detection: Report of the national research council committee to review the scientific evidence on the polygraph, 2002.

[7] Vrij, A.: Detecting lies and deceit: The psychology of lying and the implications for professional practice. John Wiley, 2000.

[8] Cestaro, V.L.: A comparison between decision accuracy rates obtained using the polygraph instrument and the computer voice stress analyzer (cvsa) in the absence of jeopardy. Technical report, DTIC Document, 1995.

[9] Burgoon, J.K., D.P. Twitchell, M.L. Jensen, T.O. Meservy, M. Adkins, J. Kruse, A.V. Deokar, G. Tsechpenakis, S. Lu, D.N. Metaxas, et al.: Detecting concealment of intent in transportation screening: a proof of concept. Intelligent Transportation Systems, IEEE Transactions on, 10(1):103-112, 2009.

[10] Ekman, P.: Telling lies: Clues to deceit in the marketplace, marriage, and politics, 1985.

[11] Bond Jr, C.F. and W.E. Fahey: False suspicion and the misperception of deceit. British Journal of Social Psychology, 26(1):41-46, 1987.

[12] Burgoon, J., J. Blair, T. Qin, and J. Nunamaker: Detecting deception through linguistic analysis. Intelligence and Security Informatics, pages 958, 2003.

[13] Bhaskaran, N., I. Nwogu, M.G. Frank, and V. Govindaraju: Lie to me: Deceit detection via online behavioral learning. In Automatic Face & Gesture Recognition and Workshops (FG 2011), 2011 IEEE International Conference on, pages 24-29. IEEE.

[14] Tsiamyrtzis, P., J. Dowdall, D. Shastri, IT Pavlidis, MG Frank, and P. Ekman: Imaging facial physiology for the detection of deceit. International Journal of Computer Vision, 71(2):197-214, 2007.

[15] Nwogu, I., M. Frank, and V. Govindaraju: An automated process for deceit detection.In Proceedings of SPIE, volume 7667, page 76670R, 2010.

[16] Meservy, T.O., M.L. Jensen, W.J. Kruse, J.K. Burgoon, and J.F. Nunamaker: Automatic extraction of deceptive behavioral cues from video. Terrorism Informatics, pages 495-516, 2008.

[17] Buller, D.B. and J.K. Burgoon: Interpersonal deception theory. Communication Theory, 6(3):203-242, 1996.

[18] Duda, R.O., P.E. Hart, D.G. Stork, et al.: Pattern classi_cation, volume 2. wiley New York, 2001.

[19] The moment of truth. 2008. http://www.imdb.com/title/tt1166919/.

[20] The moment of truth (u.s. game show). 2008. http://en.wikipedia.org/wiki/The_Moment_of_Truth_%28U.S._game_show%29.

[21] Orozco, J., F. Roca, and J. Gonzalez: Deterministic and stochastic methods for gaze tracking in real-time. In Computer Analysis of Images and Patterns, pages 45-52.Springer, 2007.

[22] Orozco, J., F.X. Roca, and J. Gonzalez: Real-time gaze tracking with appearance-based models. Machine Vision and Applications, 20(6):353-364, 2009.

[23] Frank, M.G. and P. Ekman: The ability to detect deceit generalizes across different types of high-stake lies. Journal of Personality and Social Psychology, 72(6):1429, 1997.

[24] Ketchen, D.J. and C.L. Shook: The application of cluster analysis in strategic management research: an analysis and critique. Strategic management journal, 17(6):441-458, 1996.

[25] Vrij, A. and S. Mann: Telling and detecting lies in a high-stake situation: The case of a convicted murderer. Applied Cognitive Psychology, 15(2):187-203, 2001.