

Towards Detecting Deception using K-Nearest Neighbour Model

Owolafe Otasowie
Cybersecurity Science Dept
FUTA, Nigeria

Alese Bonifacem K.
Cybersecurity Science Dept
FUTA, Nigeria

Adewale Olumide S.
Computer Science Dept
FUTA, Nigeria

ABSTRACT

Security over the years remains a major concern of all especially the law enforcement agencies. One way of arresting this concern is to be able to reliably detecting deception. Detecting deception remains a difficult task as no perfect method has been found for the detection. Past researches made use of a single cue (verbal or nonverbal), it was found that examining combinations of cues will detect deception better than examining a single cue. Since no single verbal or nonverbal cue is able to successfully detect deception the research proposes to use both the verbal and nonverbal cues to detect deception. Therefore, this research aims to develop a KNN model for classifying the extracted verbal, nonverbal and VerbNon features as deceptive or truthful. The system extracted desired features from the dataset of Perez-Rosas. The verbal cues capture the speech of the suspect while the nonverbal cues capture the facial expressions of the suspect. The verbal cues include the voice pitch (in terms of variations), frequency perturbation also known as jitters, pauses (voice or silent), and speechrate (is defined as the rate at which the suspect is speaking). The Praat (a tool for speech analysis) was used in extracting all the verbal cues. The nonverbal features were extracted using the Active Shape Model (ASM). The work was implemented in 2015a MatLab. The classification was done using KNN model. KNN performed well with VerbNon dataset with a percentage score of 96.2%.

General Terms

Machine Learning, Deception Detection, Artificial Intelligence

Keywords

VerbNon, Praat, ASM, KNN, Deception

1. INTRODUCTION

Deception, an everyday occurrence is as old as life itself. It had its origin in the Garden of Eden when Eve was deceived by the serpent. It is asserted that life and deception are inseparable since life is linked intrinsically with information which in turn is synonymous with communication. In warfare, deception dated back to 800 BC when it was first used as key to success [1]. In the biological world, some organisms display deceptive characteristics (as a way of life) like the mimicry of plants to attract pollinators or the camouflage of fish to escape predators. Some others exhibit deception as a cognitively conscious character, as in the behaviour of monkeys and of humans to mislead their colleagues to obtain a benefit for themselves [2].

The crucial question is to which behaviour attention should be paid. This question is difficult to answer, as research has shown that deception itself is not related to a unique pattern of

specific behaviours ([3]; [4]; [5]; [6]). In other words, there is nothing like Pinocchio's growing nose. However, liars might experience emotions while lying and the three most common types of emotion associated with deceit as captured in [7] are fear, excitement ('duping delight') and guilt. Detecting deception remains a difficult task [8] as no perfect method has been found for the detection [9]. In fact, multiple studies have established that lie detection results in a 50/50 chance even for experienced investigators. Although detecting deception remains difficult, investigators increase the odds for success by learning a few basic nonverbal (psychological) and verbal (speech) cues of deception. Lying requires the deceiver to keep the fact straight, make the story believable, and be able to withstand scrutiny. In [8], it was stated that when individuals tell the truth, they often make every effort to ensure that other people understand while liars on the other hand attempt to manage peoples' perceptions. Consequently, people unwittingly signal deception via nonverbal and verbal cues (Cues are those indicators or variables that can be observed and measured and are believed to be indicative of deception).

Repeated studies have shown that traditional methods of detecting deception during interviews succeed only 50% of the time, even for experienced law enforcement officers. In spite of this, investigators still need the ability to test the veracity of those they interview. To do so, investigators require a model that incorporates research with empirical experience to differentiate honesty from deception.

Unfortunately, no particular nonverbal or verbal cue evinces deception. Investigators' abilities to detect deceptive behaviour depend largely on their ability to observe, catalogue, and differentiate human behaviour. In [10], the authors stated that deception could be detected by observing non-verbal behaviour such as body language and vocal pitch. In [11], the author found that examining multiple cues was significantly more reliable indicator of deception than examining a single cue.

Since no single verbal or non-verbal cues is able to successfully detect deception (based on past works), the research proposes to use combination of verbal and nonverbal cues (VerbNon) to detect deception.

2. RELATED WORKS

Detecting deception has been an issue in scientific research as no single cue can reliably detect deception [5]. Human investigators perform a little better than chance and as such a reliable means to effectively detect deception becomes paramount.

The authors in [9] studied the behaviour of people when they are lying compared with when they are telling the truth. Their research results show that liars are less forthcoming than truth tellers, and they tell less compelling tales. The researcher also

reported that liars make a more negative impression and are tense. However, behaviours showed no discernible links, or only weak links, to deceit. Cues to deception were more pronounced when people were motivated to succeed, especially when the motivations were identity relevant rather than monetary or material. Cues to deception were also stronger when lies were about transgressions. These cues are verbal and nonverbal. Verbal cues are linguistic patterns exhibited in spoken messages while nonverbal cues are leakages or deformations that occur in the body channels of the deceiver.

The work of [12] examined certain systematically identifiable segments — called CRITICAL SEGMENTS — that bear propositional content directly related to the topics of most interest in the interrogation. They augmented the approach with techniques for adjusting the class imbalance in the data. The results, as much as 23.8% relative improvement over chance, substantially exceed human performance at the task of TRUTH and LIE classification. Further, models generated using these segments employ features consistent with hypotheses in the literature and the expectations of practitioners [13] about cues to deception.

[14] stated that since some cues (micro expressions) appears in a microseconds, detecting deception by trained and untrained professionals becomes a little better than chance. They stated further that creating or developing an automated tool that will help in flagging these deceptive cues is paramount.

[15] focused on how the behaviour of previously unseen persons can be charted using back-propagation neural network. The work was carried out using a simulated theft scenario where 15 participants were asked to either steal or not to steal some money and were later interviewed about the location of the money. A video of each interview was presented to an automatic system, which collected vectors containing nonverbal behavioural data. Each vector represented a participant's nonverbal behaviour related to "deception" or "truth" for a short period of time. These vectors were used for training and testing a back-propagation ANN which was subsequently used for charting the behavioural state of others.

In [16] the authors in their work address the question pertaining to the nature of deception language. The research aimed at the exploration of deceit in Spanish written communication. The work designed an automatic classifier based on Support Vector Machines (SVM) for the identification of deception in an *ad hoc* opinion corpus. In order to test the effectiveness of the LIWC2001 categories in Spanish, the authors drew a comparison with a Bag-of-Words (BoW) model. The results indicate that the classification of the texts was successful. They concluded that the findings were potentially applicable to forensic linguistics and opinion mining, where extensive research on languages other than English is needed.

The authors in [17] developed and implemented a system for automatically identifying deceptive and truthful statements in narratives and transcribed interviews. The research focused exclusively on verbal cues to deception for this initial experiment, ignoring at present potential prosodic cues. The authors describe a language-based analysis of deception that we was constructed and tested using "real world" sources such as criminal narratives, police interrogations and legal testimony. The analysis comprises two components: a set of deception indicators that were used for tagging a document

and an interpreter that associates tag clusters with deception likelihood. The researchers tested the analysis by identifying propositions in the corpus that could be verified as true or false and then comparing the predictions of our model against this corpus of ground truth. The analysis achieved an accuracy of 74.9%.

[11] examined the hypotheses that (1) a systematic analysis of nonverbal behaviour could be useful in the detection of deceit and (2) that lie detection would be most accurate if both verbal and nonverbal indicators of deception are taken into account. Seventy-three nursing students participated in their study about "telling lies" and either told the truth or lied about a film they had just seen. The interviews were videotaped and audiotaped, and the nonverbal behaviour (NVB) and speech content of the liars and truth tellers were analyzed, the latter with the Criteria-Based Content Analysis technique (CBCA) and the Reality Monitoring (RM) technique. Results revealed several nonverbal and verbal indicators of deception. On the basis of nonverbal behaviour alone, 78% of the lies and truths were correctly classified. The researcher speculated that a higher percentage could be correctly classified when all three detection techniques (i.e., NVB, CBCA, RM) were taken into account.

Since no single cue can reliably detect deception, a combination of verbal and nonverbal cues will help in detecting deception to a reasonably degree. [18] presented a multimodal deception detection model using real-life occurrences of deceit. They introduced a novel dataset covering recordings from public real trials and street interviews, and used this dataset to perform both qualitative and quantitative experiments. The analysis of nonverbal behaviours occurring in deceptive and truthful videos brought insight into the gestures that play a role in deception. They built classifiers relying on individual or combined sets of verbal and nonverbal features and achieve accuracies in the range of 77-82%. Their automatic system outperforms the human detection of deceit by 6-15%. Their work is the first to automatically detect instances of deceit using both verbal and nonverbal features extracted from real deception data. But the work is not a fully automated deception detection system, since they used human coders to extract the cues.

3. SYSTEM DESIGN

3.1 Dataset

The dataset consists of real-life trial videos (Perez-Rosas, 2015), some of these videos are publicly available on YouTube channels and other public websites. The dataset contains statements made by exonerees after exoneration and a few statements from defendants during crime-related TV episodes. The speakers in the videos are either defendants or witnesses. The video clips are labelled as deceptive or truthful based on a guilty verdict, not-guilty verdict, and exoneration.

The dataset consists of 121 videos including 61 deceptive and 60 truthful trial clips. The average length of the videos in the dataset is 28.0 seconds. The average video length is 27.7 seconds and 28.3 seconds for the deceptive and truthful clips, respectively. The data consists of 21 different female and 35 different male speakers, with their ages ranging between 16 and 60 years.

3.2 Feature Extraction

Features are the characteristics of the objects of interest or salient features in an image. Feature extraction is the technique of extracting these salient features from images of

different abnormal categories in such a way that class similarity is either minimized or maximized.

In classification problem, the use of salient features is essential for accuracy. The use of a model that can fit the shape of the image of interest from the dataset becomes paramount. The verbal and the nonverbal cues data was extracted from a sufficient voice and video database. The verbal Cues were extracted using Praat while the nonverbal cues were extracted using Active Shape Model (ASM).

3.3 Model Design

In designing the model relevant features were extracted from the video to form the verbal and nonverbal datasets. The features extracted from the voice repository are pitch, jitter, pause and speechrate. For extracting the Pitch feature, Praat uses the autocorrelation algorithm as shown in equation 1.

$$r_x(\tau) \approx r_{x,w}(\tau)/r_w(\tau) \quad 1$$

Where $r_x(\tau)$ represent autocorrelation of the original signal, $r_{x,w}(\tau)$ is the autocorrelation of the windowed signal and $r_w(\tau)$ is the autocorrelation of the window.

For the jitter extraction, the algorithm used is presented in equation 2.

$$jitter = \frac{100}{(N-1)\bar{\alpha}} \sum_{i=2}^N |\alpha_i - \alpha_{i-1}| \quad 2$$

α_i is the fundamental frequency

The pause will be extracted using equation 3.

$$P_a = T_t - P_t \quad 3$$

Where P_a is the total number of Pauses, T_t is Total length of time taken for the suspect to talk, P_t is the phonetic time (actual time taken to talk).

The speechrate is extracted using equation 4.

$$S_r = N_s/T_t \quad 4$$

Where speechrate is denoted as S_r , number of syllabus as N_s , and total time taken as T_t .

The nonverbal cues (that is the facial expressions) to be extracted using the Active Shape Model (ASM) are: Eyelid Blinking, Lip movement, eyebrow movement. To form the shape model, lot of training examples (in this case, different faces) were collected and the correspondence for each of the training examples were formed. Consider a person (a face), j from the set of training examples, the n^{th} feature points of the person j and for all the training set is given by equation 5.

$$z^j = \{(x_1^j, y_1^j), (x_2^j, y_2^j), (x_3^j, y_3^j) \dots (x_n^j, y_n^j)\} \quad 5$$

Since all the shapes may not be properly aligned, the shapes are rotated and translated to be centred at the origin (0, 0). After translation, the dimension of the set of aligned shapes is reduced using PCA. Any shape can then be approximated using equation 6.

$$Z = \mu + Pb \quad 6$$

where b is the model parameters, $P = V_1, V_2 \dots V_k$.

3.4 K- Nearest Neighbour

In KNN classification, neighbours closest to the data with unknown classification are used to classify the data. In using KNN, three things are required:

- a. The set of stored records known as the training dataset
- b. Distance metric to compute distance between records
- c. The value of K which is the number of nearest neighbours to retrieve.

The dataset were the features extracted using ASM and Praat. The dataset are grouped into:

- i. The verbal dataset
- ii. The nonverbal dataset
- iii. VerbNon dataset (combination of the verbal and nonverbal dataset)

Euclidean Distance: After acquiring the dataset, the next step is to calculate the distance to determine the nearest point to the new data. The Euclidean distance was used as shown in equation 7.

$$d(p, q) = \sqrt{\sum_{i=1}^8 (p_i - q_i)^2} \quad 7$$

Where (p, q) are points in the feature set.

After calculating the distance, the next step is to determine the class from the nearest neighbour list. This is done by considering the majority vote of class labels among the K -nearest neighbours.

The vote is weighed according to distance using equation 8.

$$w = \frac{1}{d^2} \quad 8$$

the value of k was calculated using equation 9.

$$k = \sqrt{D_s} \quad 9$$

Where k is the nearest neighbours to retrieve, D_s is the total number of training dataset that is considered

4. RESULT AND DISCUSSION

4.1 Performance Metrics

The metrics used for carrying out the performance evaluation are listed as:

1. False Positive Rate (FPR):

$$F_{N,R} = \frac{F_P}{T_P + F_P} \quad 10$$

2. True Positive Rate (TPR):

$$T_{P,R} = \frac{T_P}{T_P + F_N} \quad 11$$

3. Accuracy: The overall accuracy is given by the sum of true and false utterances correctly classified, out of all the classifications carried out. It is the number of correct predictions over the total number of predictions.

$$Accuracy = \sum \left(\frac{T_P + F_P}{T_P + T_N + F_P + F_N} \right) \quad 12$$

Where T_P, T_N, F_P and F_N are the True positive, True negative, False positive and False negative values respectively. Figure 3 shows the accuracy and the error recorded for the research.

Confusion Matrix: It is a table used to describe the performance of the classification model on the dataset.

Figure 1 shows the confusion matrix for K- Nearest Neighbour classifier. From the figure, 833 data were correctly classified as truthful which corresponds to 96.2% while 17 were wrongly classified as deceptive corresponding to 3.2%. Also, 516 were correctly classified as deceptive representing 96.8% and 33 falsely classified as truthful representing 3.8%. Figure 2 Shows the ROC of the KNN.

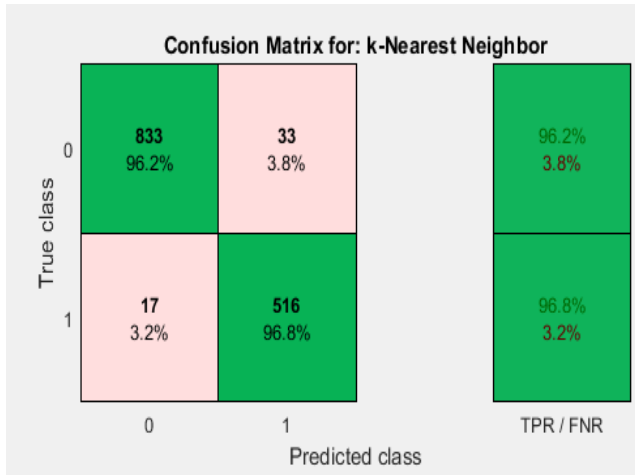


Figure 1: Confusion Matrix

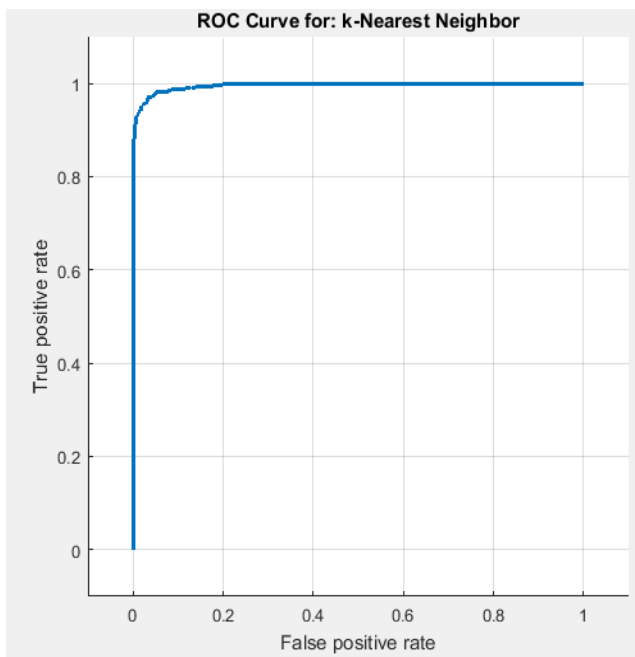


Figure 2: ROC Curve

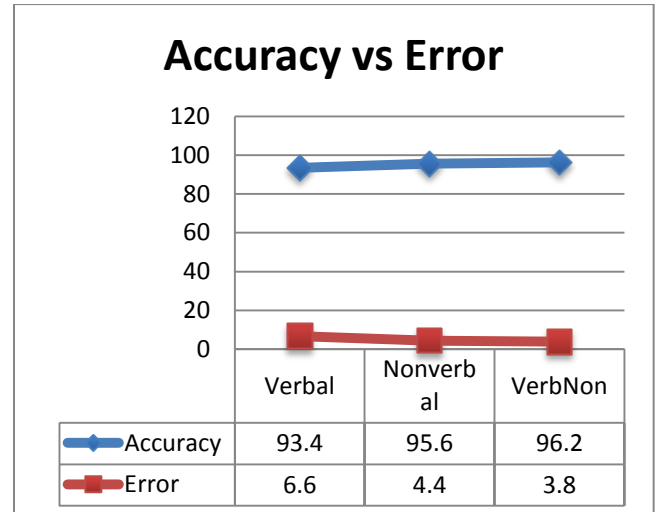


Figure 3: Accuracy vs Error

From Figure 3, it is observed that as Accuracy is increasing error is reducing. KNN performs well with combination of verbal and nonverbal cues.

This research uses verbal, nonverbal and VerbNon cues (a combination of both verbal and nonverbal cues) to detect deception. The system was implemented using Matlab 2015a on window 7 with 2GB RAM. Each of the extracted dataset was divided into training data and test data. The classification was done using KNN model on the different dataset and at the end of the comparative analysis it was discovered that KNN model work well on VerbNon dataset to detect deception. The result obtained using only verbal cue was 93.4% while that of nonverbal cue was 95.6% but on VerbNon yielded 96.2% which is far better than the chance level of 50%.

5. CONCLUSION

Deception detection is an involved social issue because to successfully deceive, the deceiver has to formulate a story that is internally consistent while hiding emotions and true intentions. Facial expressions and voice play a critical role in the identification of deception as shown in this research. Previous research made use of only one cue but this research made use of verbal, nonverbal and VerbNon cues. The developed system was able to perform better than chance and trained professionals with a difference of 45.1%. For future research, we hope to use Decision Tree and Support Vector Machine.

6. REFERENCES

- [1] Baron-Cohen, Simon, Wheelwright, Sally, Hill, Jacqueline, Raste, Yogini and Plumb, Ian. 2001. The Reading the Mind in the Eyes Test Revision Version: A Study with Normal Adults, and Adults with Asperger Syndrome or High-functioning Autism. *Journal of Child Psychology and Psychiatry*, 241-251, vol.42, no 2.
- [2] Picornell I.C. 2013. Cues to Deception in a Textual Narrative Context: Lying in Written Witness Statements. A Thesis Submitted for the Degree of Doctor of Philosophy to Centre for Forensic Linguistics, School of Languages and Social Sciences, Aston University.
- [3] Depaulo B. M., Stone J. I., and Lassiter G. D. 1985. Deceiving and Detecting Deceit. *The Self and Social*. 323-370. New York: McGraw-Hill,

- [4] Ekman Paul. 1992. An Argument for Basic Emotions. *Journal of Cognition and Emotion*, 169-200, vol 6, no 3/4.
- [5] Vrij A. 2000. Telling and detecting lies as a function of raising the stakes. In *New Trends in Criminal Investigation and Evidence*, by M.M Kommer, J.F Nijboer and J.M Reintjes C.M Breur, 699-709. Belgium: Antwerpen.
- [6] Zuckermzn M., Depaulo B. M., and Rosenthal R. 1981. Verbal and Nonverbal Communication of Deception. In *Advances in Experimental Social Psychology*, by Berkowitz L., 1-59. New York: Academic Press.
- [7] Ekman P., 1985. *Telling Lies: Clues to deceit in the market place, politics and marriage*. New York: Norton.
- [8] Navarro J., and Schafer R. J. 2001. *Detecting Deception*. FBI Law Enforcement Bulletin.
- [9] Depaulo B. M., Linsay J. J., Malone B. E., Muhlenbruck L., Charlton K., and Cooper H. 2003. Cues to Deception. *Psychological Bulletin*. 74-112, vol 129.
- [10] Sporer S. L., and Schwandt B. 2006. Paraverbal Correlates of Deception: A meta-analysis. *Applied Cognitive Psychology*. 421-446, vol. 20.
- [11] Vrij A. 2008. *Detecting Lies and Deceit: Pitfalls and Opportunities*. Chichester: Wiley.
- [12] Enos F., Benus S., Cautin R. L., Graciarena M., Hirschberg J., and Shriberg E. 2006. Personality factors in human deception detection: Comparing human to machine performance." *Proceedings of the 9th International Conference on Spoken Language Processing, ISCA 2006*. Pittsburgh, USA.
- [13] Reid J. and Associates. 2000. *The Reid Technique of Interviewing and INterrogation*. Chicago: John E. Reid and Associates.
- [14] Zhou L., Twitchell D. P., Qin T., Burgoon J. K., and Nunamaker J. F., Jr. 2003. An exploratory study into deception detection in text-based computer-mediated communication. *Proceedings of the 36th Annual Hawaii International Conference on System Sciences* . Hawaii, 2003.
- [15] Rothwell J., Bandar Z., O'Shea J., and McLean D. 2006. Charting the behavioural state of a person using a Backpropagation Neural Network. *Journal of Neural Computing and Applications*.
- [16] Almela A., Valencia-Garcia R., and Cantos P. 2012. Seeing through Deception: A Computational approach to deceit detection in written communication." In *proceedings of the Workshop on Computational Approaches to Deception Detection*. Avignon, France. 15-22.
- [17] Bachenko Joan, Eileen Fitzpatrick and Michael Schonwetter. 2008. Verification and implementation of language based deception indicators in civil and criminal narratives. *proceedings of the 22nd International Conference on Computational Linguistics*. Manchester, U.K. 41-48.
- [18]