Structural Optimization of Respiration, EDA and Cardio Activity Using a Genetic AI

Raymond Nelson

Abstract

This project involved the use of a balanced sample of n=36 field polygraph exams and a simple genetic algorithm to compute a weighting function for polygraph signals that would optimize the classification of deception and truth-telling. A genetic algorithm is a simple form of machine-learning that can be used to address complex problems in optimization, classification, search, and other data-analytic contexts. EDA accounted for, or explained, 54% of the diagnostic variance in the sample data. Cardiovascular activity accounted for 34% of the difference variance in the guilty and innocent sample sampling data. The weighting coefficient for respiration was 12%. This weighting function is somewhat similar to other weighting functions in the polygraph literature. Although this study contributes little additional information to the published knowledge base, in addition to being computationally intensive and involving a small sample size, results of this study demonstrate the potential use for advanced computing techniques in polygraph research. Computing technology is more abundant and less expensive than in the past. Continued interest is indicated for both weighted EDA solutions, and the use of computational machine learning methods in polygraph research.

Polygraph testing, although often referred to conveniently as a lie detector, does not detect or measure lies, but instead relies on data that is primarily autonomic. These include respiration movement, electrodermal activity, cardiovascular activity and sometimes vasomotor activity. Analysis of polygraph data involves a series of functions similar to other data analytic contexts, including feature extraction, numerical transformation and data reduction, the use of some form of likelihood function, and structured decision rules to parse a categorical test result from the numerical and probabilistic data. An important challenge of any multivariate analysis is the calculation, or optimization, of a statistical function that specifies an optimal combination of the different sources of data that will achieve a desired objective.

Optimization refers to the calculation or computation of a best attainable solution. Optimization is a data-analytic approach to solution finding, as opposed to solution-finding through conjecture or anecdotal example, subjective opinion, or even expert opinion (equivalent to subjective opinion and conjecture). One way to determine the optimal structural combination of sensor data will be to test every possible combination. However, attempting to test every possible solution will be an expensive and time-consuming expedition. The number of possible weighting coefficients or structural combinations of respiration, EDA, and cardiovascular is potentially infinite. To gain insight into the possibilities, if weighting coefficients are regarded as normalized decimal proportions (summing to 1) there are 166,650 possible combination using only two decimals of precision. With the addition of a fourth recording sensor (i.e., vasomotor), the number of possible structural functions will be 4,082,925. Three decimals of precision would increase the possible combinations exponentially, though with potentially little benefit.

Another method to optimize the structural combination of respiration, EDA, and cardiovascular activity (or any combination of response features) would be to use traditional statistical methods such as linear discriminate analysis, linear regression or logistic regression. A more modern approach to optimization and classification problems (also search problems, and prediction problems) is to use statistical learning theory (Hastie, Tibshirani & Friedman, 2009; James, Witten, Hastie & Tibshirani, 2013;), also referred to as machine-learning (ML) and artificial-intelligence (AI).

An important difference between AI and the traditional statistical approach is that the traditional approach involves a researcher or scientist who develops a hypothesis (possible solution) about a possible answer to a research question. The researcher then designs an experiment to falsify the hypothesis or to compare the hypothesis and null hypothesis to determine which is more consistent with the observed data. The AI approach allows a computing machine to both suggest and test numerous possible hypotheses. Thus, the machine is said to "learn" a solution from its experience with the data.

This project involved the use of a genetic algorithm (Goldberg, 1989; Mitchel, 1996) to compute structural combinations of respiration, EDA and cardiovascular activity. The optimization question is this: what is the best structural weighting for data from each of the polygraph sensors? In this context, *best* is defined as achieves the greatest number of correct decisions when classifying the sample cases as deceptive or truthful.

Data

Data consisted of a small sample of n=18 confirmed deceptive and n=18 confirmed truthful polygraph cases. The sample cases were conducted with a diagnostic polygraph format with two relevant questions. Cases were conducted by a large metropolitan police agency, consisted of respiration, EDA and cardiovascular activity data, and were confirmed through a combination of confession and extra-polygraphic evidence. Examinees were criminal suspects who authorized the examination, including the use of the data in anonymous form for research, program evaluation instruction and quality control. All exams consisted of three iterations (three charts) of the sequence of the test questions. All examinations consisted of sensors for thoracic and abdominal respiration movement, EDA, cardiovascular activity and an activity sensor.

The two relevant question diagnostic format is used for event-specific diagnostic polygraphs. It includes two relevant questions and three comparison questions, along with other procedural questions. When using the two relevant question diagnostic polygraph format, each relevant question is evaluated with the preceding or subsequent comparison question depending on which comparison question has produced the greater change in physiological activity. All exams were conducted and recorded using the Lafayette LX4000 polygraph instrument.

Data were exported from the proprietary binary file format to the NCCA ASCII text format using a data sampling rate of 30 samples per second. Data were then imported to the R Language and Environment for Statistical Computing (R Core Team, 2019) for analysis. All feature extraction, numerical transformation, data reduction, likelihood calculations and decision rules were executed automatically in the R computing environment. The respiratory feature of interest was the reduction of respiration activity in response to the test stimuli, associated with attempts to conceal one's deception. The EDA feature of interest was the change in y-axis value from an onset of a positive slope segment to the peak of reaction, associated with increased activity in the sympathetic division of the autonomic nervous system. For cardiovascular activity data the feature of interest was the change in y-axis value, also associated with relative blood pressure and activity in the autonomic nervous system.

Feature extraction was performed for each sensor for each relevant question (RQ) and each comparison question (CQ). Respiration data was measured as the mean of respiration line excursion (RLE; the absolute difference of each subsequent respiration sample) for a one-second moving average from stimulus onset to 15 seconds post stimulus onset excluding the data from one second before to one second after the recorded verbal answer. This measurement is thought to be more robust against distortions at the point of verbal answer and is not influenced by the length of the 15 second evaluation window – effects with different measurement periods will have a similar metric. EDA reactions were measured as the onset of a positive slope segment during a response onset window (ROW) from .5 seconds after stimulus onset to 5 seconds after the verbal answer to the greatest y-axis (vertical) distance to subsequent peak of reaction (onset of negative slope) within evaluation window (EW) from stimulus onset to 15 seconds after stimulus onset. If there was no response onset during the ROW a response onset was inferred statistically during positive slope segments using a z-test of the variance of one second mean difference of each subsequent EDA sample. A response onset was imputed if the difference in variance for a two, one-second windows exceeded the alpha = .001 boundary. This can be visualized as a substantial increase in positive slope angle within a positive slope segment during the ROW. Cardiovascular activity was extracted by first calculating the mean of all cardio sensor samples.

This can be thought of, and plotted, as the mid-line between the systolic and diastolic peaks. Cardiovascular activity changes were then extracted, using the cardio mid-line, using a procedure similar to the one for the EDA data.

All measurement values were dimensionless. That is, they were not indexed to any physical quantity, SI unit , or derived measurement value. Dimensionless values were then transformed to objective ordinal rank values using a three-point coding scheme [-1, 0, +1] familiar to field polygraph examiners. For each of the recording sensors, extracted values for each presentation of each RQ was compared to the preceding or subsequent CQ depending on which CQ produced the greater change in physiological activity. Scores were coded as +1 if the change in physiological activity was greater at the CQ and were coded as -1 if the change in physiological activity was greater at the RQ. Tied values (tied ranks) were coded as 0. For EDA and cardiovascular activity, a greater extracted value was indicative of a greater change in physiology. However, because the respiratory feature of interest involved the reduction of respiration activity, sign values were inverted so that smaller extracted values were interpreted as a greater change in physiological activity.

Non-parametric rank values were then reduced to subtotal scores for each RQ through summation. Subtotal scores were then summed to achieve a grand total score for each exam. The analytic theory of the polygraph test postulates that greater changes in physiological activity are loaded at different types of test stimuli as a function of deception and truth-telling in response to relevant target stimuli (Nelson, 2015, 2016). Under this theory, grand total scores of this type can be expected to be greater than zero for innocent examinees and less than zero for guilty examinees. The genetic algorithm was used to determine the weighting coefficients that can be assigned to scores from each of the recording sensors to maximize the number of correct classifications.

Analysis

A genetic algorithm can be thought of as a Monte Carlo method, involving the use of random numbers to create numerous possible solutions to a question or analytic problem. [See Eckhardt (1987), Metropolis, (1987), and Metropolis and Ulam (1949) for more information on Monte Carlo methods]. A genetic algorithm consists of simple rules such as the following:

1. Creation of numerous (say, m=1000) random possible solutions for the structural weighting of respiration, EDA and cardiovascular activity data,

2. Testing the effects of each possible solution with all of the sample cases,

3. Survival of the best solutions (natural selection) – discard the 50% that performs weakest and keep the 50% that achieves the best classification,

1 International System of Units (French: Système international d'unités, abreviated as SI). SI base units include the following: the meter as a measurement of length or distance, the kilogram as a unit of mass, the second as a unit of time, the ampere as a unit of electric current, the kelvin as a unit of temperature, the candela as a unit for luminosity, and the mole as a unit for the quantity of a substance. All other measurement units are derived from these SI base units. Measurement of any quantity requires both a physical quantity to measure and a defined unit of measurement

4. Split each of the surviving solutions into two parts and randomly connect them (recombination) to make a new iteration of m possible solutions for the structural weighting of the sensor data – now informed by the previous experience,

5. Introduce random variation (mutation) to a small portion of the new solutions – to potentially find better solutions that were not included in the previous solutions,

6. Repeat steps 2-5 a large number of times,

7. Stop at some point – either after a specified number of iterations (say, 30,000), or in response to the achievement of a stated objective (e.g. a desired level of accuracy), or when the structural model stops improving, and finally,

8. Choose the structural solution that achieves the greatest effect size

Results

The genetic algorithm used objective integer-level rank order input data and produced the weighting function shown in Table 1. EDA accounted for or explained over half of the diagnostic variance in the sample data. Cardiovascular activity accounted for approximately one-third of the difference between the guilty and innocent sample sampling data. Respiration data explained slightly over 10% of the diagnostic variance. This weighting function is somewhat similar to other weighting functions in the polygraph literature, including the discriminate function reported by Nelson, Krapohl and Handler (2008) in the development of the Objective Scoring System Version-3, also shown in Table 1.

Discussion

This project involved the use of a balanced sample of n=36 field polygraph exams and a simple genetic algorithm to compute a weighting function for polygraph signals that would optimize the classification of deception and truth-telling. A genetic algorithm, and other ML techniques, can achieve a very close approximation of an optimal solution with only a few thousand (sometimes many thousand) iterations. Response features in this study were coded with an objective rank method using positive and negative values [-1, 0, +1] by comparing responses to relevant and comparison stimuli. Input data were intentionally naive as to the relative importance of the data from different recording sensors, and the algorithm output is a weighting function that will optimize the diagnostic variance of the extracted data.

Table 1. Weighting function.

Sensor	Normalized weighting function	
	Genetic Algorithm	OSS-3
Respiration activity	.12	.19
EDA	.54	.53
Cardiovascular activity	.34	.28

EDA data accounted for over 50% of the variance while cardiovascular data accounted for approximately 1/3 of the diagnostic variance. Respiration data accounted for the smallest portion of diagnostic variance. This weighting coefficients are similar to other published information. Some manual scoring protocols approximate this weighting function by doubling EDA scores.

The procedures in this study differ from those commonly used scoring in field polygraph programs, in which manual/visual feature extraction continues to be a dominant method for the interpretation of polygraph test data. It also differs from most studies on automated algorithm development in its use of ordinal integer-level numerical coding. Results from this study add additional confirmation to existing knowledge on the relative importance of polygraph signals, and may be helpful to better understand polygraph scoring methods such methods such as the OSS (Krapohl, 2002; Krapohl & McManus, 1999) and ESS (Nelson, Krapohl & Handler, 2008; Nelson et al., 2011; Nelson 2017).

Limitations of this project include the small sample size, and the limited information available about the case confirmation. Despite the sample size, results from this study appear to be consistent with other information on the structural weighting of polygraph signals. Another, limitation of this project, related to the small sample size, is the absence of a hold-out sample. No attempt was made, during this project, to test the effectiveness of the weighting function with other data. Also, no attempt was made to test the effectiveness of the weighting function with the study input data, as doing so would incur a risk of over-fitting a conclusion with the small input sample, and thereby overestimating its effectiveness. Another potential limitation, related to the use of Monte Carlo methods with small sample sizes, is that replication of these results may be subject to both sampling variation and Monte Carlo variation. This limitation is mitigated by the results of other studies on signal weighting in manual scoring methods – such as those already cited, the one by Nelson and Handler (2018) – that demonstrate the effects of weighting the EDA data more than the other sensor data. A final limitation of this study is that it is computationally intensive. However, computing power is much more abundant and much less expensive than in the past. Thoughtful use of computing and analytic technologies can help to improve and advance the science and field practice of polygraphic credibility assessment testing.

In consideration of the volume of existing information, results of this study are not surprising, and the results of this study contribute little new knowledge to the science and field practice of polygraph testing. Optimization of respiration, EDA and cardiovascular activity has previously been demonstrated using a variety of methods, including logistic regression and discriminate analysis and other methods. Monte Carlo methods have been described in previous polygraph studies. These results are interesting because they serve to add further confirmation of extant knowledge regarding polygraph signals, and it introduces and demonstrates the potential use of ML/AI techniques in polygraph studies. Continued interest is indicated for both weighted EDA solutions, and the use of computational machine learning methods in polygraph research.

References

Eckhardt, R. (1987). *Stan Ulam, John von Neumann, and the Monte Carlo method.* Los Alamos Science (15), 131–137.

Goldberg, D. (1989). *Genetic Algorithms in Search, Optimization and Machine Learning.* Addison-Wesley Professional.

James, G., Witten, D., Hastie, T. & Tibshirani, R. (2013). *An Introduction to Statistical Learning: with applications in R.* Springer.

Krapohl, D. J. (2002). Short report: Update for the objective scoring system. *Polygraph,* 31, 298-302.

Krapohl, D. & McManus, B. (1999). An objective method for manually scoring polygraph data. *Polygraph*, 28, 209-222.

Kroese, D. P., Brereton, T., Taimre, T., & Botev, Z. I. (2014). Why the Monte Carlo method is so important today. *Wiley Interdisciplinary Reviews: Computational Statistics* (6), 386–92.

Metropolis, N. (1987). The beginning of the Monte Carlo method. Los Alamos Science (*1987 Special Issue dedicated to Stanislaw Ulam*), 125–130.

Metropolis, N., Ulam, S. (1949). The Monte Carlo Method. *Journal of the American Statistical Association,* 44(247), 335–341.

Mitchell, M. (1996). *An Introduction to Genetic Algorithms.* Cambridge, MA: MIT Press.

Nelson, R. (2015). Scientific basis for polygraph testing. *Polygraph 41(1)*, 21-61.

Nelson, R. (2016). Scientific (analytic) theory of polygraph testing. *APA Magazine 49(5)*, 69-82.

Nelson, R. (2017). Multinomial reference distributions for the Empirical Scoring System. Polygraph & Forensic Credibility Assessment, 46 (2). 81-115.

Nelson, R. & Handler, M. (2018). Reducing inconclusive results: a descriptive analysis of decision rules, weighted electrodermal scores and multinomial cut-scores. *Polygraph & Forensic Credibility Assessment, 47 (2)* 108-121.

Nelson, R., Krapohl, D., & Handler, M. (2008). Brute force comparison: A Monte Carlo study of the Objective Scoring System version 3 (OSS-3) and human polygraph scorers. *Polygraph, 37,* 185-215.

R Core Team (2019). R: *A language and environment for statistical computing.* R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/.