

Optimización Estructural la Actividad Respiratoria, del EDA y Cardiovascular Utilizando IA Genética

Raymond Nelson

Abstract

Este proyecto involucró el uso de una muestra equilibrada de $n=36$ exámenes de polígrafo de campo y de un algoritmo genético simple para calcular una función de ponderación de señales poligráficas que optimizaría la clasificación del engaño y la veracidad. Un algoritmo genético es una forma simple de aprendizaje automatizado que se puede utilizar para abordar problemas complejos de optimización, clasificación, búsqueda y de otros contextos analíticos de datos. EDA representó, o explicó, el 54% de la varianza diagnóstica en los datos de la muestra. La actividad cardiovascular dio cuenta del 34% de la diferencia de varianza en los datos de muestreo de la muestra de culpables e inocentes. El coeficiente de ponderación para la respiración fue del 12%. Esta función de ponderación es de alguna manera similar a otras funciones de ponderación en la literatura poligráfica. Aunque este estudio aporta poca información adicional a la base de conocimiento publicada, además de ser computacionalmente intensivo y haber involucrado un pequeño tamaño de muestra, los resultados de este estudio demuestran el uso potencial de técnicas informáticas avanzadas en la investigación poligráfica. La tecnología informática es más abundante y menos costosa que en el pasado. Se indica un interés continuo tanto por las soluciones ponderadas del EDA como por el uso de métodos computacionales de aprendizaje automatizado en la investigación poligráfica.

Las pruebas de polígrafo, aunque a menudo se refieren convenientemente como un detector de mentiras, no detectan ni miden mentiras, sino que se basan en datos que son principalmente autónomos. Estos incluyen movimientos respiratorios, la actividad electrodérmica y la actividad cardiovascular, y algunos sistemas poligráficos también hacen uso de la actividad vasomotora. El análisis de los datos del polígrafo involucra una serie de funciones similares a los de otros contextos analíticos de datos, que incluyen extracción de características, transformación numérica y reducción de datos, uso de alguna forma de función de probabilidad y reglas de decisión estructuradas para analizar un resultado de prueba categórico a partir de los datos numéricos y probabilísticos. Un reto importante en cualquier análisis multivariado es el cálculo u optimización de una función estadística que especifique una combinación óptima de las diferentes fuentes de datos que lograrán el objetivo deseado.

La American Polygraph Association (APA) tiene los derechos de autor de este artículo, y aparece aquí con el permiso de la APA.

Traductor. rodolfo@poligrafia.com.mx

Polygraph & Forensic Credibility Assessment, 2019, 48 (2)

La optimización se refiere al cálculo o cómputo de la mejor solución posible. La optimización es un enfoque del análisis de datos para la búsqueda de soluciones, en oposición a la búsqueda de soluciones a través de conjeturas o ejemplos anecdóticos, opiniones subjetivas o incluso opiniones de expertos (equivalente a la opinión subjetiva y conjetura). Una forma de determinar la combinación estructural óptima de los datos del sensor será probar cada una de las combinaciones posibles. Sin embargo, intentar probar todas las soluciones posibles será una expedición larga en tiempo y costosa. El número de posibles coeficientes de ponderación o combinaciones estructurales de respiración, EDA y cardiovascular es potencialmente infinito. Para tener una idea de las posibilidades, si los coeficientes de ponderación se consideran proporciones decimales normalizadas (sumándose a 1), hay 166,650 combinaciones posibles utilizando solo dos decimales de precisión. Con la adición de un cuarto sensor (registro vasomotor), el número de funciones estructurales posibles será de 4,082,925. Tres decimales de precisión aumentarían las posibles combinaciones exponencialmente, aunque con un beneficio potencialmente pequeño.

Otro método para optimizar la combinación estructural de la actividad respiratoria, EDA y cardiovascular (o cualquier combinación de características de respuesta) será utilizar métodos estadísticos tradicionales, como el análisis discriminatorio lineal, la regresión lineal o la regresión logística. Un enfoque moderno para los problemas de optimización y clasificación (también en problemas de búsqueda y de predicción) es utilizar la teoría del aprendizaje estadístico (Hastie, Tibshirani y Friedman, 2009; James, Witten, Hastie y Tibshirani, 2013;), también conocida como aprendizaje automatizado (*en Inglés Machine Learning ML*) e inteligencia artificial (IA).

Una diferencia importante entre la IA y la estadística tradicional es que el enfoque tradicional involucra a un investigador o científico que desarrolla una hipótesis (posible solución) sobre una posible respuesta a una pregunta de investigación. Luego, el investigador diseña un experimento para falsear la hipótesis o comparar la hipótesis contra la hipótesis nula para determinar cuál es más consistente con los datos observados. El enfoque de la IA permite que una máquina de cómputo sugiera y pruebe numerosas hipótesis posibles. Por lo tanto, se dice que la máquina "aprende" una solución de su experiencia con los datos.

Este proyecto involucró el uso de un algoritmo genético (Goldberg, 1989; Mitchel, 1996) para hacer el cómputo de la combinación estructural de la actividad respiratoria, EDA y cardiovascular. La pregunta de optimización es la siguiente: ¿cuál es la mejor ponderación estructural para los datos de cada uno de los sensores poligráficos? En este contexto, se define como el que logra el mayor número de decisiones correctas al clasificar los casos de muestra como de engaño o veraces.

Datos

Los datos consistieron en una pequeña muestra de n=18 casos poligráficos confirmados como engaño y n=18 casos confirmados veraces. Los casos de muestra se realizaron con un formato de polígrafo de diagnóstico con dos preguntas relevantes. Los casos fueron aplicados por una policía metropolitana muy grande, que constaban de datos de actividad respiratoria, EDA y cardiovascular, y que se confirmaron mediante una combinación de confesión y evidencia extra poligráfica. Los examinados eran sospechosos de delitos que autorizaron el examen, incluyendo el uso de los datos en forma anónima para investigación, programas de instrucción de evaluación y control de calidad. Todos los exámenes consistieron en tres iteraciones (tres gráficas) de la secuencia de preguntas de prueba. Todos los exámenes constaron en sensores de actividad respiratoria torácica y abdominal, EDA y cardiovascular, además de un sensor de actividad.

El formato diagnóstico de dos preguntas relevantes se utiliza para los polígrafos de diagnóstico de eventos específicos e incluye dos preguntas relevantes y tres preguntas de comparación, junto con otras preguntas de procedimiento. Cuando se utiliza el formato de polígrafo de diagnóstico de dos preguntas relevantes, cada pregunta relevante se evalúa contra la pregunta de comparación anterior o posterior dependiendo de qué pregunta de comparación haya producido el mayor cambio en la actividad fisiológica. Todos los exámenes se realizaron y registraron utilizando el instrumento poligráfico Lafayette LX4000.

Los datos se exportaron del formato patentado de archivo binario al formato de texto ASCII de NCCA utilizando una tasa de muestreo de datos de 30 muestras por segundo. Luego, los datos se importaron al lenguaje y entorno R para computación estadística (R Core Team, 2019) para su análisis. Todas las funciones de extracción, transformación numérica, reducción de datos, cálculos de probabilidad y reglas de decisión se ejecutaron automáticamente en el entorno informático R. La característica respiratoria de interés fue la reducción de la actividad respiratoria en respuesta a los estímulos de prueba, asociada con los intentos de ocultar el engaño. La característica EDA de interés fue el cambio en el valor del *eje-y* desde el inicio de un segmento de pendiente positiva hasta el pico de reacción, asociado con una mayor actividad en la división simpática del sistema nervioso autónomo. Para los datos de actividad cardiovascular, la característica de interés fue el cambio en el valor del *eje-y*, también asociado con la presión sanguínea relativa y la actividad en el sistema nervioso autónomo.

La extracción de características se realizó para cada sensor, para cada pregunta relevante (RQ) y para cada pregunta de comparación (CQ). Los datos de respiración se

midieron como la media de la excursión de la línea de respiración (RLE; la diferencia absoluta de cada muestra de respiración posterior) para un promedio de movimiento de un segundo desde el inicio del estímulo hasta 15 segundos después del inicio del estímulo, excluyendo los datos de un segundo antes a un segundo después de la respuesta verbal registrada. Esta medición se piensa es más robusta en contra de las distorsiones en el punto de respuesta verbal y no se ve influenciada por la duración de la ventana de evaluación de 15 segundos - los efectos con diferentes periodos de medición tendrán una métrica similar. Las reacciones del EDA se midieron como el inicio de un segmento de pendiente positiva durante una ventana de inicio de respuesta (ROW) desde .5 segundos después del inicio del estímulo hasta 5 segundos después de la respuesta verbal a la mayor distancia del *eje-y* (vertical) al pico de reacción subsecuente (inicio de pendiente negativa) dentro de la ventana de evaluación (EW) desde el inicio del estímulo hasta 15 segundos después del inicio del estímulo. Si no había inicio de respuesta durante el ROW, se infirió estadísticamente un inicio de respuesta durante los segmentos de pendiente positiva utilizando una prueba z de la varianza de la diferencia media de un segundo de cada muestra subsecuente del EDA. Un inicio de respuesta era imputado si la diferencia en la varianza para dos ventanas de un segundo excedía el límite alfa = .001. Esto se puede visualizar como un aumento sustancial en el ángulo de pendiente positivo dentro de un segmento de pendiente positivo durante el ROW. La actividad cardiovascular se extrajo calculando primero la media de todas las muestras del sensor de cardio.

Esto puede pensarse y trazarse como la línea media entre los picos sistólicos y diastólicos. Luego se extrajeron los cambios en la actividad cardiovascular, usando la línea media de cardio, usando un procedimiento similar al de los datos de EDA.

Todos los valores de medición fueron adimensionales. Es decir, no se indexaron a ninguna cantidad física, unidad SI¹, o valor de medición derivado. Los valores adimensionales se transformaron luego en valores de rango ordinal objetivo utilizando un esquema de codificación de tres puntos [-1, 0, +1] común para los examinadores de polígrafo de campo. Para cada uno de los sensores de registro, los valores extraídos para cada presentación de cada RQ se compararon con la CQ anterior o posterior, dependiendo de qué CQ produjera el mayor cambio en la actividad fisiológica.

1 Sistema Internacional de Unidades (Francés: *Système international d'unités*, abreviado como SI). Las unidades base SI incluyen las siguientes: el metro como una medida de longitud o distancia, el kilogramo como una unidad de masa, el segundo como una unidad de tiempo, el amperio como unidad de corriente eléctrica, el kelvin como unidad de temperatura, la candela como unidad de luminosidad y el mole como unidad para la cantidad de una sustancia. Todas las demás unidades de medida se derivan de estas unidades base SI. La medición de cualquier cantidad requiere tanto de una cantidad física para medir como de una unidad de medida definida.

Las puntuaciones se codificaron como +1 si el cambio en la actividad fisiológica fue mayor en la CQ, y se codificaron como -1 si el cambio en la actividad fisiológica fue mayor en la RQ. Los valores empatados (rangos empatados) se codificaron como 0. Para el EDA y la actividad cardiovascular, un mayor valor extraído era indicativo de un mayor cambio en la fisiología. Sin embargo, debido a que la característica respiratoria de interés implicaba la reducción de la actividad respiratoria, los valores de los signos se invirtieron para que los valores extraídos más pequeños se interpretaran como un cambio mayor en la actividad fisiológica.

Los valores de rango no paramétricos se redujeron a puntuaciones subtotales para cada RQ mediante sumatoria. Los puntajes subtotales se sumaron para lograr un puntaje de gran total para cada examen. La teoría analítica de la prueba del polígrafo postula que los mayores cambios en la actividad fisiológica se cargan ante diferentes tipos de estímulos de prueba en función del engaño y la veracidad, en respuesta a los estímulos relevantes objetivo (Nelson, 2015, 2016). Según esta teoría, se puede esperar que las puntuaciones de gran total de este tipo sean mayores que cero para los examinados inocentes y menores que cero para los examinados culpables. El algoritmo genético se utilizó para determinar los coeficientes de ponderación que se pueden asignar a las puntuaciones de cada uno de los sensores de registro para maximizar el número de clasificaciones correctas.

Análisis

Un algoritmo genético puede considerarse como un método de Monte Carlo, que involucra el uso de números aleatorios para crear numerosas soluciones posibles a una pregunta o problema analítico. [Ver Eckhardt (1987), Metrópolis, (1987) y Metrópolis y Ulam (1949) para más información sobre los métodos de Monte Carlo]. Un algoritmo genético consiste en reglas simples como las siguientes:

1. Creación de numerosas (por ejemplo, $m = 1000$) soluciones posibles aleatorias para la ponderación estructural de los datos de actividad respiratoria, EDA y cardiovascular,
2. Probar los efectos de cada posible solución con todos los casos de muestra,
3. Supervivencia de las mejores soluciones (selección natural): descarte el 50% que tenga un desempeño más débil y conserve el 50% que logre la mejor clasificación,
4. Separación de cada una de las soluciones supervivientes en dos partes y conexión aleatoria (recombinación) para hacer una nueva iteración de m

soluciones posibles para la ponderación estructural de los datos del sensor - ahora informada por la experiencia previa.

5. Introducción de una variación aleatoria (mutación) a una pequeña porción de nuevas soluciones, para encontrar mejores soluciones potenciales que no se incluían en las soluciones anteriores,
6. Repetición de los pasos 2-5 una gran cantidad de veces,
7. Detención en algún punto - ya sea después de un número específico de iteraciones (por ejemplo, 30,000), o en respuesta al logro de un objetivo establecido (por ejemplo, un nivel deseado de precisión), o cuando el modelo estructural deja de mejorar, y finalmente
8. Elección de la solución estructural que logra el mayor tamaño de efecto.

Resultados

El algoritmo genético utilizó datos de entrada de orden de rango de nivel entero objetivo y produjo la función de ponderación que se muestra en la Tabla 1. El EDA representó o explicó más de la mitad de la varianza de diagnóstico de los datos de la muestra. La actividad cardiovascular representó aproximadamente un tercio de la diferencia entre los datos de muestreo de la muestra de culpables e inocentes. Los datos de respiración explicaron un poco más del 10% de la varianza diagnóstica. Esta función de ponderación es algo similar a otras funciones de ponderación en la literatura poligráfica, incluyendo la función de discriminación reportada por Nelson, Krapohl y Handler (2008) en el desarrollo del Sistema de Puntuación Objetiva (OSS) Versión-3, que también se muestra en la Tabla 1.

Tabla 1. Función de ponderación.

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

Discusión

Este proyecto involucró el uso de una muestra equilibrada de $n=36$ exámenes de polígrafo de campo y un algoritmo genético simple para calcular una función de ponderación de señales poligráficas que optimizaría la clasificación del engaño y veracidad. Un algoritmo genético y otras técnicas de ML pueden lograr una aproximación muy cercana de una solución óptima con solo unos pocos miles (a veces muchos miles) de iteraciones. Las características de respuesta en este estudio se codificaron con un método de rango objetivo utilizando valores positivos y negativos [-1, 0, +1] comparando las respuestas ante estímulos relevantes y de comparación. Los datos de entrada fueron intencionalmente ingenuos con respecto a la importancia relativa de los datos de los diferentes sensores de registro, y la salida del algoritmo es una función de ponderación que optimizará la varianza de diagnóstico de los datos extraídos.

Los datos de EDA representaron más del 50% de la varianza, mientras que los datos cardiovasculares representaron aproximadamente $1/3$ de la varianza diagnóstica. Los datos de respiración representaron la porción más pequeña de la varianza diagnóstica. Estos coeficientes de ponderación son similares a otra información publicada. Algunos protocolos de puntuación manual aproximan esta función de ponderación al duplicar las puntuaciones de EDA.

Los procedimientos en este estudio difieren de los puntajes comúnmente utilizados en los programas de polígrafo de campo, en los cuales la extracción manual/visual de características sigue siendo un método dominante para la interpretación de los datos de prueba de polígrafo. También difiere de la mayoría de los estudios sobre el desarrollo automatizado de algoritmos en su uso de la codificación numérica de nivel entero ordinal. Los resultados de este estudio agregan confirmación adicional al conocimiento existente sobre la importancia relativa de las señales poligráficas, y pueden ser útiles para comprender mejor los métodos de puntuación poligráfica, tales como el OSS (Krapohl, 2002; Krapohl & McManus, 1999) y ESS (Nelson, Krapohl y Handler, 2008; Nelson et al., 2011; Nelson 2017).

Las limitaciones de este proyecto incluyen el pequeño tamaño de la muestra y la información limitada disponible sobre la confirmación de los casos. A pesar del tamaño de la muestra, los resultados de este estudio parecen ser consistentes con otra información sobre la ponderación estructural de las señales poligráficas. Otra limitación de este proyecto, relacionada con el pequeño tamaño de la muestra, es la ausencia de una muestra reservada. No se hizo ningún intento, durante este proyecto, de probar la efectividad de la función de ponderación con otros datos. Además, no se hizo ningún intento para probar la efectividad de la función de ponderación con los

datos de entrada del estudio, ya que al hacerlo se correría el riesgo de sobre ajustar una conclusión con la pequeña muestra de entrada y, por lo tanto, sobreestimar su efectividad. Otra limitación potencial, relacionada con el uso de métodos de Monte Carlo con tamaños de muestra pequeños, es que la réplica de estos resultados puede estar sujeta tanto a la variación de muestreo como a la variación de Monte Carlo. Esta limitación se ve mitigada por los resultados de otros estudios sobre la ponderación de la señal en los métodos de puntuación manual - como los ya citados, el de Nelson y Handler (2018)- que demuestran los efectos de ponderar los datos EDA sobre de los datos de otros sensores. Una limitación final de este estudio es que es computacionalmente intensivo. Sin embargo, la potencia informática es mucho más abundante y mucho menos costosa que en el pasado. El uso cuidadoso de las tecnologías informáticas y analíticas puede ayudar a mejorar y avanzar la ciencia y la práctica de campo de las pruebas de evaluación de credibilidad poligráfica.

Teniendo en cuenta el volumen de información existente, los resultados de este estudio no son sorprendentes, y los resultados de este estudio aportan poco conocimiento nuevo a la ciencia y a la práctica de campo de las pruebas de polígrafo. La optimización de la actividad respiratoria, EDA y cardiovascular se ha demostrado previamente utilizando una variedad de métodos, incluida la regresión logística, el análisis discriminador y otros métodos. Los métodos de Monte Carlo se han descrito en estudios poligráficos anteriores. Estos resultados son interesantes porque sirven para agregar confirmación del conocimiento existente con respecto a las señales de polígrafo, e introduce y demuestra el uso potencial de las técnicas de ML / IA en los estudios de poligrafía. Se indica un interés continuo tanto por las soluciones EDA ponderadas como por el uso de métodos computacionales de aprendizaje automático en la investigación poligráfica.

Referencias

- 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/>.