

“Big data” y psiquiatría

Natalie C. Soto

*Instituto de Efectividad Clínica y Sanitaria (IECS), Buenos Aires, Argentina
E-mail: nsoto@iecs.org.ar*

Resumen

La popularidad del término “big data” se ha incrementado exponencialmente en los últimos años. Este término se utiliza habitualmente para hacer referencia tanto a conjuntos de datos demasiado grandes o complejos para ser abordados con estrategias tradicionales, así como a los métodos de ciencia de datos utilizados para analizarlos. Esta revisión sintetiza la literatura sobre los usos implementados o potenciales de “big data” en psiquiatría. La literatura identificada es muy diversa y abarca un espectro que va desde la utilización de ciencia de datos para predicción de eventos, como conducta suicida o episodios psicóticos, hasta el surgimiento de un nuevo campo interdisciplinario denominado psiquiatría computacional.

Palabras clave: “Big data” - Psiquiatría - Ciencia de datos - “Data mining” - Psiquiatría computacional.

BIG DATA AND PSYCHIATRY

Abstract

The popularity of the term “big data” has increased exponentially in recent years. This term is commonly used to refer both to data sets that are too large or complex to be approached with traditional strategies, and to the data science methods used to analyze them. This review synthesizes the literature on the implemented or potential uses of big data in psychiatry. The identified literature is very diverse, covering a spectrum ranging from the use of data science methods for the prediction of events such as suicidal behavior or psychotic episodes to the emergence of a new interdisciplinary field called computational psychiatry.

Key words: Big data - Psychiatry - Data science - Data mining - Computational psychiatry.

Introducción

El término “big data” se ha vuelto muy popular en los últimos años, apareciendo en un número creciente de publicaciones científicas (1). No existe una única manera de definirlo. En líneas generales es un término que da cuenta de los desafíos inherentes al procesamiento de grandes cantidades de datos estructurados y no estructurados, y se utiliza con frecuencia para hacer referencia tanto a los conjuntos de datos en sí mismos como a los métodos de ciencia de datos utilizados para analizarlos.

El objetivo del presente artículo es presentar una revisión narrativa de los usos implementados o potenciales de conjuntos de datos caracterizables como “big data” y los métodos de la ciencia de datos en el campo de la salud mental. No pretende ser una revisión exhaustiva, sino reflejar el espectro de oportunidades que se están explorando.

Métodos

Se realizó una búsqueda de literatura científica en bases de datos específicas (PubMed y Lilacs) y motores de búsqueda generales (Google) utilizando los términos “mental health”, “psychiatry”, “big data”, “machine learning”, “artificial intelligence”, “data science”, “data mining” y “computational psychiatry”. Se realizaron consultas con expertos que sugirieron términos adicionales para la estrategia de búsqueda y publicaciones específicas.

En una primera etapa se revisaron todos los resultados de la búsqueda por título y resumen, y en una segunda etapa por texto completo, seleccionándose sólo aquellos que cumplieran los criterios de inclusión.

Dado que el foco de la revisión es presentar una síntesis cualitativa de la literatura, se incluyeron revisiones narrativas o sistemáticas sobre la utilización de conjuntos de datos caracterizables como “big data” o que utilizaran metodología de ciencia de datos, y estudios que aplicaran conjuntos de datos o abordajes no considerados en las revisiones.

Resultados

Se identificaron 46 artículos que cumplían los criterios de inclusión. Para su mejor descripción, la literatura identificada se clasificó en dos grupos: por un lado, los artículos que reflejan el uso de conjuntos de datos o metodologías de análisis con un objetivo clínico-epidemiológico tradicional, y por el otro, la literatura que describe y da ejemplos del nuevo campo de la psiquiatría computacional.

Implementaciones con objetivos clínicos/epidemiológicos tradicionales

Se identificaron 19 artículos que dan cuenta de implementaciones clínicas o epidemiológicas, que se presentan a continuación según el origen de los datos (2-20).

Historia Clínica Electrónica

La historia clínica electrónica (HCE) puede calificarse como “big data” dado que contiene grandes volúmenes de información que se acumulan tanto en campos estructurados como en campos de texto libre (4,9). Los datos estructurados tales como la edad, el sexo, el diagnóstico y las fechas asociadas al uso de servicios sanitarios (admisiones, altas, etc.), están fácilmente disponibles para su uso en investigación, y son relativamente fáciles de anonimizar. Sin embargo, su validez y precisión pueden ser inferiores a la información contenida en campos no estructurados. El texto libre es menos accesible para el análisis y se anonimiza con menos facilidad; sin embargo, la información contenida en el texto es potencialmente la más valiosa para la investigación. Para poder analizar esta información se han desarrollado e implementado metodologías de ciencia de datos para análisis de texto y se ha visto que la información detectada a través de estas metodologías puede, por ejemplo, lograr una sensibilidad mayor que aquella basada en datos estructurados para la detección de pacientes de patologías específicas. Otro uso propuesto de los datos de la HCE es el desarrollo de algoritmos que generen retroalimentación en tiempo real de información para la toma de decisiones sobre situaciones clínicas, como el riesgo de suicidio.

Los datos de la HCE de atención primaria se han utilizado también para investigar la salud física de las personas con enfermedades mentales graves o frecuentes.

Registros específicos

Los registros son bases de datos especializadas de seguimiento de personas en contacto con el sistema de salud mental o de los propios servicios (9). Son un recurso útil para investigar patrones de uso de servicios y sus determinantes individuales y sociales. Estos registros pueden relevar datos de distinto tipo, por ejemplo, una base de datos en los Países Bajos registra episodios de aislamiento y contención física, buscando información para impulsar la mejora del servicio.

Bases de datos administrativas

Las bases de datos administrativas han sido utilizadas para extraer conclusiones sobre el uso de los servicios sanitarios a través de diagnósticos en registros de egresos hospitalarios o facturación de procedimientos o prescripciones (9). Esta información tiende a ser utilizada para describir la prevalencia de trastornos tratados, patrones de prescripción, y comparar los tratamientos administrados en el “mundo real” con las recomendaciones. Cuando los datos incluyen tanto prescripciones como diagnóstico de eventos, también pueden ser útiles en farmacovigilancia, para investigar eventos adversos, como diferencias en la seguridad de diferentes antipsicóticos en pacientes mayores. Las bases de datos administrativas también se pueden usar para describir los

costos de los tratamientos y los programas de atención, entre otros.

Enlace de registros

Las bases de datos mencionadas pueden adquirir nuevas dimensiones cuando se vinculan con los datos de otras fuentes a nivel del individuo (9). Esto se facilita en países con un número de identificación único para sus residentes: por ejemplo, los investigadores que utilizan los registros en algunos países de población pueden relacionar los informes de muertes por suicidio con los registros de diagnósticos psiquiátricos y médicos, períodos de ausencia por enfermedad del trabajo y datos de desempleo. Algunos registros permiten asociar el dato también a los padres y hermanos y se han utilizado para investigar los posibles factores de riesgo para la esquizofrenia.

Encuestas, biobancos y repositorios internacionales de datos

Las encuestas, biobancos y repositorios de datos de neuroimágenes son recursos utilizados en estudios realizados tanto en adultos como en niños (8,9,17). Algunos ejemplos incluyen el programa Carga Global de la Enfermedad de la OMS utiliza encuestas estandarizadas de salud mental (basadas en la Entrevista Diagnóstica Compuesta-CIDI) llevadas a cabo periódicamente por los equipos de investigación locales en los países miembros, encuestas poblacionales de representatividad nacional, biobancos, como UK Biobank del Reino Unido, con datos genéticos y datos de fenotipo de salud mental a través de un cuestionario en línea basado en el formulario corto de la entrevista CIDI (9).

Estudios con recolección de datos ad hoc para predicción de suicidio y psicosis

Dos publicaciones reportaron el uso de datos de seguimiento longitudinal con encuestas semi-estructuradas analizadas con estrategias de aprendizaje automático para la predicción de riesgo de suicidio (7, 10).

Uno de los estudios más mencionados en la literatura sobre aplicación de ciencia de datos en salud mental, es el publicado en 2015 por Bedi y cols., seguido de un estudio de validación publicado recientemente (18,19). En estos estudios, el análisis de las características semánticas y sintácticas del lenguaje en jóvenes con alto riesgo de psicosis permitió predecir psicosis con mayor precisión que la que se derivó de entrevistas clínicas.

Redes sociales y foros online

Dos revisiones publicadas en 2016 y 2017 (Conway y cols. y Wongkoblapp y cols.) relevaron la utilización de datos de redes sociales y estrategias de ciencias de datos (aprendizaje automático y procesamiento del lenguaje natural) (6,16). Se reportó que fueron utilizados para seguimiento de indicadores poblacionales (como preva-

lencia de depresión), para la detección de indicadores de riesgo de determinados trastornos (como características de las publicaciones en redes sociales antes del parto para predecir depresión post-parto) y para identificar características del lenguaje asociadas a trastornos específicos.

En el marco de una competencia convocada en un workshop de lingüística computacional y psicología clínica, investigadores de distintos países desarrollaron un método para detectar de forma automática, en foros de salud mental, aquellos mensajes que requirieran la intervención de un moderador por ser indicadores de riesgo (20).

Historiales de búsqueda en internet

Una revisión del año 2017 relevó estudios que utilizaron registros de búsquedas en internet en el contexto de proyectos de psicología social (12). Entre los desenlaces evaluados se incluyeron proyectos sobre conducta suicida y trastornos mentales. Esta revisión identificó estudios que evaluaban la correlación de las búsquedas con prevalencia poblacional de determinados trastornos, así como la capacidad de predecir muertes por suicidio.

Un estudio exploró la posibilidad de detectar predictores de un conjunto de patologías (algunas de ellas trastornos mentales) a través de la detección de patrones en sus historiales de búsqueda en internet (5).

Aplicaciones web desarrolladas para personalización de intervenciones psicoterapéuticas

Se identificaron cuatro publicaciones en las que se utilizaron estrategias de ciencia de datos para personalizar intervenciones terapéuticas a partir de metodologías específicas de recopilación de datos de los pacientes, en combinación con metodologías de ciencias de datos (aprendizaje automático) (2,3,11,15).

Datos recolectados por sensores de teléfonos celulares y dispositivos

Mohr y cols. relevaron en 2017 la literatura sobre investigación en salud mental basada en el uso de sensores personales, como aquellos que se encuentran disponibles en la mayoría de los teléfonos celulares (13). En los estudios identificados, los sensores fueron utilizados para registrar marcadores comportamentales (como sueño, contexto social, estado de ánimo y estrés) y para explorar la posibilidad de detectar trastornos mentales o su severidad a través de la utilización de los marcadores mencionados.

Literatura sobre psiquiatría computacional

La psiquiatría computacional es uno de los emergentes más importantes de las grandes cantidades de datos y poder computacional que se han vuelto disponibles en los últimos años para los profesionales de salud mental. Se identificaron 27 artículos que presentaban revisiones

y ejemplos de implementación de estudios en psiquiatría computacional (21-47). Una definición propuesta en la literatura identificada propone que la psiquiatría computacional es un campo interdisciplinario que sostiene que la experiencia humana, la toma de decisiones y el comportamiento, pueden entenderse en términos de cómo construimos representaciones mentales de los estados del mundo y actuamos para influir en esos estados lo más útilmente posible (22). Bajo esta mirada, la enfermedad psiquiátrica se conceptualiza en términos de fallas para lograr esta interacción óptima. Lo que se entiende por óptimo podría establecerse por los estados y valores del individuo y el estado del mundo. Este estado del mundo incluye otras mentes y representaciones mentales, así como estructuras sociales complejas. En este enfoque, la intervención sería posible en múltiples niveles; se podría cambiar el entorno de las personas con problemas mentales (tal vez ayudándolas a encontrar vivienda o empleo), modificar la representación de parámetros y usar esas representaciones modificadas para guiar su comportamiento de una manera óptima según sus valores (por ejemplo, mediante intervenciones psicoterapéuticas), o corregir alteraciones neurobiológicas (por ejemplo, con medicación). La psiquiatría computacional se orienta hacia una mirada biopsicosocial ya que cada nivel de análisis (a través de neuronas, circuitos, procesos cognitivos e interacciones sociales) sólo puede comprenderse completamente al caracterizar su asociación con los otros niveles. El análisis desde esta perspectiva se basa en el uso de modelos generativos que es una descripción probabilística de cómo las causas de alto nivel generan datos de bajo nivel. Saber cómo las causas generan datos permite que un modelo genere datos sintéticos o simulados a partir de causas determinadas. Esta descripción generativa puede ser de cómo la actividad cerebral genera datos de imágenes cerebrales, o de cómo los estados en el mundo evolucionan y afectan la toma de decisiones de una persona; en este

último caso, comúnmente utilizado en este campo, se “modela el modelo” del mundo del cerebro. Las estadísticas Bayesianas y las técnicas de aprendizaje automático (ambas enmarcadas dentro de la denominada ciencia de datos) permiten que esta descripción completa se evalúe con datos reales. Formular un modelo generativo requiere una descripción explícita de los detalles matemáticos del proceso cognitivo o neuronal. Esto obliga al investigador no sólo a pensar en lo que realmente significan las construcciones particulares (como “atención” o “relevancia”), sino también a ser explícito sobre lo que se asume y lo que se ignora.

Si bien esta perspectiva permite trascender las categorías diagnósticas, algunas revisiones explícitamente enmarcadas en esta mirada relevaban publicaciones de psiquiatría computacional en esquizofrenia, trastorno de déficit de atención e hiperactividad, trastorno límite de la personalidad y trastorno bipolar (43-47).

Reflexiones finales

La literatura identificada demuestra que los usos potenciales de los grandes datos y las metodologías de ciencia de datos son muy diversos y abarcan un amplio espectro que va desde la utilización de metodologías de ciencia de datos para predicción de eventos como suicidio o psicosis, hasta el surgimiento de un nuevo campo de investigación interdisciplinario denominado psiquiatría computacional.

Al considerar lo expuesto, debe tenerse en cuenta que no se trata de una revisión exhaustiva y que tampoco se exploraron en profundidad potenciales limitaciones metodológicas o prácticas. Por otra parte, debe recordarse que, si bien no se realizó un análisis sobre este aspecto, el uso de datos debe ir siempre acompañado de una discusión sobre las implicancias éticas que acarrea, teniendo en cuenta el origen de los mismos y el contexto de su utilización. ■

Referencias bibliográficas

1. Gu D, Li J, Li X, Liang C. Visualizing the knowledge structure and evolution of big data research in healthcare informatics. *International journal of medical informatics*. 2017;98:22-32.
2. Burns MN, Begale M, Duffecy J, et al. Harnessing context sensing to develop a mobile intervention for depression. *Journal of medical Internet research*. 2011;13(3):e55.
3. Kelly J, Gooding P, Pratt D, Ainsworth J, Welford M, Tarrier N. Intelligent real-time therapy: harnessing the power of machine learning to optimise the delivery of momentary cognitive-behavioural interventions. *Journal of mental health (Abingdon, England)*. 2012;21(4):404-414.
4. Monteith S, Glenn T, Geddes J, Bauer M. Big data are coming to psychiatry: a general introduction. *International journal of bipolar disorders*. 2015;3(1):21.
5. Yom-Tov E, Borsa D, Hayward AC, McKendry RA, Cox IJ. Automatic identification of Web-based risk markers for health events. *Journal of medical Internet research*. 2015;17(1):e29.
6. Conway M, O'Connor D. Social Media, Big Data, and Mental Health: Current Advances and Ethical Implications. *Current opinion in psychology*. 2016;9:77-82.
7. Cook BL, Progovac AM, Chen P, Mullin B, Hou S, Baca-Garcia E. Novel Use of Natural Language Processing (NLP) to Predict Suicidal Ideation and Psychiatric Symptoms in a Text-Based Mental Health Intervention in Madrid. *Computational and mathematical methods in medicine*. 2016;2016:8708434.
8. McIntosh AM, Stewart R, John A, et al. Data science for mental health: a UK perspective on a global challenge. *The lancet Psychiatry*. 2016;3(10):993-998.

9. Stewart R, Davis K. 'Big data' in mental health research: current status and emerging possibilities. *Social psychiatry and psychiatric epidemiology*. 2016;51(8):1055-1072.
10. Barros J, Morales S, Echavarrí O, et al. Suicide detection in Chile: proposing a predictive model for suicide risk in a clinical sample of patients with mood disorders. *Revista brasileira de psiquiatria (Sao Paulo, Brazil : 1999)*. 2017;39(1):1-11.
11. D'Alfonso S, Santesteban-Echarri O, Rice S, et al. Artificial Intelligence-Assisted Online Social Therapy for Youth Mental Health. *Frontiers in psychology*. 2017;8:796.
12. Lai K, Lee YX, Chen H, Yu R. Research on Web Search Behavior: How Online Query Data Inform Social Psychology. *Cyberpsychology, behavior and social networking*. 2017;20(10):596-602.
13. Mohr DC, Zhang M, Schueller SM. Personal Sensing: Understanding Mental Health Using Ubiquitous Sensors and Machine Learning. *Annual review of clinical psychology*. 2017;13:23-47.
14. Perna G, Grassi M, Caldirola D, Nemeroff CB. The revolution of personalized psychiatry: will technology make it happen sooner? *Psychological medicine*. 2017:1-9.
15. Schueller SM, Aguilera A, Mohr DC. Ecological momentary interventions for depression and anxiety. *Depression and anxiety*. 2017;34(6):540-545.
16. Wongkoblaph A, Vadillo MA, Curcin V. Researching Mental Health Disorders in the Era of Social Media: Systematic Review. *Journal of medical Internet research*. 2017;19(6):e228.
17. Blair LM. Publicly Available Data and Pediatric Mental Health: Leveraging Big Data to Answer Big Questions for Children. *Journal of pediatric health care : official publication of National Association of Pediatric Nurse Associates & Practitioners*. 2016;30(1):84-87.
18. Bedi G, Carrillo F, Cecchi GA, et al. Automated analysis of free speech predicts psychosis onset in high-risk youths. *NPJ schizophrenia*. 2015;1:15030.
19. Corcoran CM, Carrillo F, Fernandez-Slezak D, et al. Prediction of psychosis across protocols and risk cohorts using automated language analysis. *World psychiatry : official journal of the World Psychiatric Association (WPA)*. 2018;17(1):67-75.
20. Milne DN. Triaging content in online peer-support: an overview of the 2017 CLPsych shared task. 2017; <http://clpsych.org/shared-task-2017/>.
21. Montague PR, Dolan RJ, Friston KJ, Dayan P. Computational psychiatry. *Trends in cognitive sciences*. 2012;16(1):72-80.
22. Corlett PR, Fletcher PC. Computational psychiatry: a Rosetta Stone linking the brain to mental illness. *The lancet Psychiatry*. 2014;1(5):399-402.
23. Friston KJ, Stephan KE, Montague R, Dolan RJ. Computational psychiatry: the brain as a phantastic organ. *The lancet Psychiatry*. 2014;1(2):148-158.
24. Wang XJ, Krystal JH. Computational psychiatry. *Neuron*. 2014;84(3):638-654.
25. Moutoussis M, Story GW, Dolan RJ. The computational psychiatry of reward: broken brains or misguided minds? *Frontiers in psychology*. 2015;6:1445.
26. Adams RA, Huys QJ, Roiser JP. Computational Psychiatry: towards a mathematically informed understanding of mental illness. *Journal of neurology, neurosurgery, and psychiatry*. 2016;87(1):53-63.
27. Huys QJ, Maia TV, Frank MJ. Computational psychiatry as a bridge from neuroscience to clinical applications. *Nature neuroscience*. 2016;19(3):404-413.
28. Paulus MP, Huys QJ, Maia TV. A Roadmap for the Development of Applied Computational Psychiatry. *Biological psychiatry Cognitive neuroscience and neuroimaging*. 2016;1(5):386-392.
29. Teufel C, Fletcher PC. The promises and pitfalls of applying computational models to neurological and psychiatric disorders. *Brain: a journal of neurology*. 2016;139(Pt 10):2600-2608.
30. Chekroud AM, Lane CE, Ross DA. Computational Psychiatry: Embracing Uncertainty and Focusing on Individuals, Not Averages. *Biological psychiatry*. 2017;82(6):e45-e47.
31. Iglesias S, Tomiello S, Schneebeli M, Stephan KE. Models of neuromodulation for computational psychiatry. *Wiley interdisciplinary reviews Cognitive science*. 2017;8(3).
32. Lytton WW, Arle J, Bobashev G, et al. Multiscale modeling in the clinic: diseases of the brain and nervous system. *Brain informatics*. 2017;4(4):219-230.
33. Maia TV, Huys QJM, Frank MJ. Theory-Based Computational Psychiatry. *Biological psychiatry*. 2017;82(6):382-384.
34. Moutoussis M, Eldar E, Dolan RJ. Building a New Field of Computational Psychiatry. *Biological psychiatry*. 2017;82(6):388-390.
35. Petzschner FH, Weber LAE, Gard T, Stephan KE. Computational Psychosomatics and Computational Psychiatry: Toward a Joint Framework for Differential Diagnosis. *Biological psychiatry*. 2017;82(6):421-430.
36. Reiter A, Heinz A, Deserno L. Linking social context and addiction neuroscience: a computational psychiatry approach. *Nature reviews Neuroscience*. 2017;18(7):450.
37. Yahata N, Kasai K, Kawato M. Computational neuroscience approach to biomarkers and treatments for mental disorders. *Psychiatry and clinical neurosciences*. 2017;71(4):215-237.
38. Frassle S, Yao Y, Schobi D, Aponte EA, Heinzle J, Stephan KE. Generative models for clinical applications in computational psychiatry. *Wiley interdisciplinary reviews Cognitive science*. 2018.
39. Huys QJM. Advancing Clinical Improvements for Patients Using the Theory-Driven and Data-Driven Branches of Computational Psychiatry. *JAMA psychiatry*. 2018.
40. Fernandes BS, Williams LM, Steiner J, Leboyer M, Carvalho AF, Berk M. The new field of 'precision psychiatry'. *BMC medicine*. 2017;15(1):80.
41. Schwartenbeck P, Friston K. Computational Phenotyping in Psychiatry: A Worked Example. *eNeuro*. 2016;3(4).
42. Mota NB, Copelli M, Ribeiro S. Computational Tracking of Mental Health in Youth: Latin American Contributions to a Low-Cost and Effective Solution for Early Psychiatric Diagnosis. *New directions for child and adolescent development*. 2016;2016(152):59-69.
43. Librenza-Garcia D, Kotzian BJ, Yang J, et al. The impact of machine learning techniques in the study of bipolar disorder: A systematic review. *Neuroscience and biobehavioral reviews*. 2017;80:538-554.
44. Valton V, Romaniuk L, Douglas Steele J, Lawrie S, Series P. Comprehensive review: Computational modelling of schizophrenia. *Neuroscience and biobehavioral reviews*. 2017;83:631-646.
45. Hauser TU, Fiore VG, Moutoussis M, Dolan RJ. Computational Psychiatry of ADHD: Neural Gain Impairments across Marrian Levels of Analysis. *Trends in neurosciences*. 2016;39(2):63-73.
46. Krystal JH, Murray JD, Chekroud AM, et al. Computational Psychiatry and the Challenge of Schizophrenia. *Schizophrenia bulletin*. 2017;43(3):473-475.
47. Fineberg SK, Stahl D, Corlett P. Computational Psychiatry in Borderline Personality Disorder. *Current behavioral neuroscience reports*. 2017;4(1):31-40.