

Estimación de los casos de COVID-19 y brotes en tiempo real a través de llamadas telefónicas



*Franco Marsico**

Resumen

Uno de los principales problemas para controlar la propagación de la epidemia de COVID-19 es el retraso en la confirmación de los casos. Tener información sobre los cambios en la evolución de la epidemia o los brotes antes de la confirmación en el laboratorio es crucial en la toma de decisiones para las políticas de Salud Pública. Presentamos un algoritmo para estimar el número de casos de COVID-19 en tiempo real utilizando los datos de las llamadas telefónicas a una línea de COVID. Modelando las llamadas como fondo (proporcional a la población), más señal (proporcional a los infectados), ajustamos el indicador en la provincia de Buenos Aires (Argentina) con un coeficiente de determinación $R^2 > 0,85$. Este resultado nos permite estimar el número de casos dado el número de llamadas de un distrito específico, días antes de que los resultados del laboratorio estén disponibles. Validamos el algoritmo con datos reales. Mostramos cómo usar el algoritmo para hacer un seguimiento de la epidemia y presentamos la alarma de brotes tempranos para detectar brotes antes de los resultados del laboratorio. Un punto clave en el algoritmo desarrollado es un seguimiento detallado de las incertidumbres en las estimaciones, ya que la alarma utiliza la importancia de los observables como indicador principal para detectar una anomalía. Presentamos los detalles del ejemplo explícito en Villa Azul (Quilmes) donde esta herramienta resultó crucial para controlar un brote a tiempo. Las

* Profesor de Introducción a Bases de Datos y Minería de Datos, Tecnicatura Informática Aplicada a la Salud, Universidad de José C Paz. E-mail: franco.lmarsico@gmail.com

herramientas presentadas han sido diseñadas con urgencia con los datos disponibles en el momento del desarrollo y, por lo tanto, tienen sus limitaciones que describimos y discutimos. Consideramos posibles mejoras en las herramientas, muchas de las cuales están actualmente en desarrollo.

Palabras clave

vigilancia epidemiológica - vigilancia molecular - sistemas de alarma temprana - COVID-19

1. Introducción

La epidemia de COVID-19 está causando daños globales a prácticamente todos los aspectos de la sociedad mundial desde principios de 2020. Aunque se está haciendo un gran esfuerzo en muchos campos de la ciencia para mitigar sus efectos, la enfermedad se está propagando continuamente y, en muchas regiones, una segunda ola está causando grandes preocupaciones. Las dificultades para controlar la epidemia se deben en parte a la combinación crucial de ser altamente contagiosa (Gao et al., 2020), tener un largo período de incubación (Lee et al., 2020) durante el cual es posible el contagio unos días antes de la aparición de los síntomas (He et al., 2020), tener casos leves o asintomáticos (Gao et al., 2020) y también porque el diagnóstico puede tardar días desde que se contacta con el sistema sanitario. En particular, esto último hace que los brotes se propaguen y la epidemia evolucione mientras se procesan los resultados de laboratorio. Este efecto es más importante en los países de ingresos bajos y medianos debido a problemas operativos y logísticos, generalmente causados por las desigualdades tecnológicas y económicas (Ahmed, 2020; Verhagen et al, 2020).

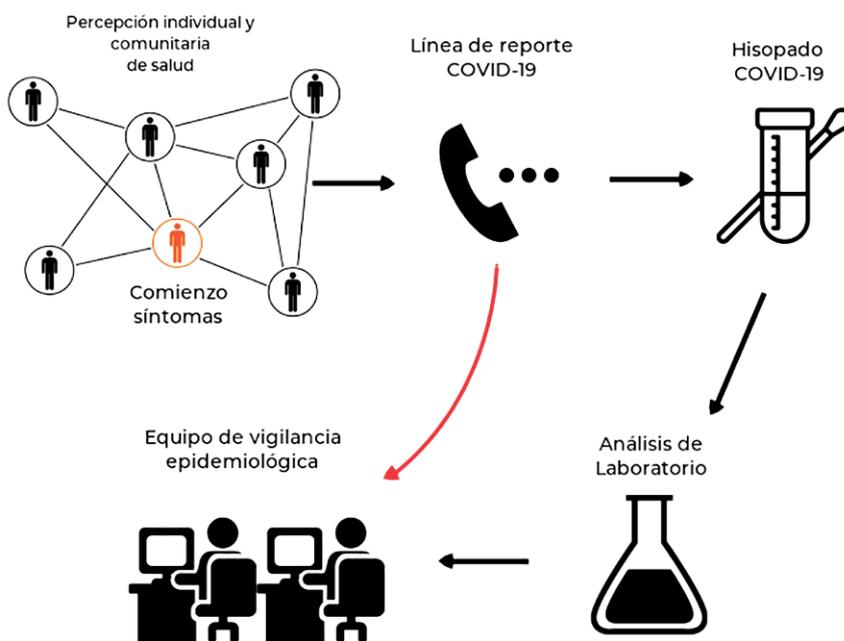
En este trabajo presentamos un método para mitigar los efectos de la epidemia mediante la estimación del número de casos de COVID-19 sin tener que esperar las confirmaciones de laboratorio. Esto proporciona al sistema de salud una herramienta para reaccionar con anticipación y evaluar las políticas de salud pública actuales o futuras.

En las catástrofes masivas los sistemas de alerta temprana (SAT) juegan un papel clave para la mitigación de daños (Texier et al., 2017; Goniewicz et al. 2019; Kyriacos et al. 2014) disminuyendo los tiempos de respuesta por parte de los equipos de emergencia y mejorando su efectividad. La principal estrategia de los SAT en la vigilancia de enfermedades infecciosas es la incorporación de la información producida lo más cercanamente posible al momento de la infección (Grinsberg et al., 2020; Krause et al., 2007). En este caso, la aparición de los síntomas y su detección por parte de la percepción de salud individual y comunitaria es la primera señal detectable de un brote. Los sistemas de alerta temprana basados en la vigilancia sindrómica se han aplicado en la vigilancia epidemiológica para la identificación y confirmación temprana de los brotes (Lombardo et al., 2003; Paylin et al., 2003; Katz et al., 2011; Stoto et al., 2004; Hope et al., 2006). Una de las principales características del sistema de alerta temprana es la utilización de la información sanitaria proporcionada por la población para activar las alarmas locales. Hoy en día, con el amplio uso de aplicaciones de telefonía celular y

líneas telefónicas específicas del sistema de salud, constantemente se generan importantes bases de datos con información sobre la vigilancia sindrómica (Diwan et al., 2015). La geolocalización juega un papel principal en la definición espacial y temporal de los brotes detectados por los sistemas de alerta temprana (Chen et al., 2011).

En la provincia de Buenos Aires (Argentina), la línea telefónica 148 de COVID-19 es uno de los primeros contactos entre una persona que se cree infectada y el sistema de salud. El equipo capacitado de Atención Médica recibe y responde a las preguntas de las personas generando, simultáneamente, una base de datos de vigilancia sindrómica. Si la persona tiene síntomas que podrían indicar una infección por COVID-19, se le indica que siga el protocolo correspondiente. Es importante señalar que esa base de datos de síndromes se utilizó como insumo para la estimación de casos y la detección de brotes en la provincia de Buenos Aires. A medida que las personas llaman a la línea COVID según su percepción de la salud, los operadores formados en COVID determinan si corresponden a casos sospechosos o de contacto cercano. En tal caso, su historial pasa al equipo de vigilancia epidemiológica y se ordena un hisopado para análisis de COVID-19. Algunos días después, el resultado del hisopado se añade al registro correspondiente. El flujo de información ordinario se esquematiza con flechas negras. El algoritmo descrito en este documento funciona con la primera parte de la información que se entrega en el flujo de trabajo a medida que los operadores determinan que el caso se corresponde con los criterios correspondientes (flecha roja) (Figura 1).

Figura 1. Flujo de trabajo de la línea 148 de COVID.



2. Métodos

Cuando la gente llama a la línea COVID 148, entra en un menú de voz automático en el que una de las opciones corresponde a síntomas similares a los de COVID. Cuando el usuario entra en esta opción, su llamada es atendida por un operador capacitado y un breve cuestionario sobre su experiencia indica si la llamada no pasa el umbral para ser registrada o si corresponde a una de las dos categorías registradas: contacto cercano y caso sospechoso. Si la llamada corresponde a alguna de estas categorías, el operador registra sus datos y, en particular, el distrito desde el que llama. En la Figura 1 se muestra el flujo de trabajo de la línea COVID. En la fase inicial de la implantación del sistema, el registro no incluía información fiable sobre la dirección exacta del usuario. Este hecho crucial nos llevó a desarrollar el sistema que explicamos a continuación, restringiendo nuestra información sobre el usuario solo a su distrito. Aunque las futuras actualizaciones del sistema están proporcionando una localización más precisa de la llamada, el trabajo actual se restringe al distrito del llamante y solo una vez que su llamada es tomada por un operador entrenado por COVID.

Presentamos el modelo matemático para estimar los nuevos infectados usando los datos de la llamada telefónica, y lo aplicamos a PBA. El razonamiento en esta sección sigue las mismas líneas que en la Ref. (Álvarez y Marsico, 2018), pero con diferentes propósitos y diferentes filtros en el conjunto de datos. Consideramos un conjunto de datos de llamadas de muchos distritos y durante un rango de tiempo determinado a una línea COVID. Cada una de estas llamadas puede ser

fondo: personas con síntomas similares pero no infectadas

señal: personas infectadas con COVID-19.

En este caso el “*fondo*” se corresponde a aquellas llamadas no atribuibles a casos reales de COVID-19 y que por lo tanto pueden ser caracterizadas como “*ruido*”. En cambio, la “*señal*” es lo que realmente se quiere medir/detectar y se corresponde a las llamadas relacionadas con casos de COVID-19.

Bajo supuestos razonables de homogeneidad en el espacio y el tiempo podemos modelar que las llamadas de fondo en cada distrito y ventana de tiempo son proporcionales a la población total del distrito y a la duración de la ventana de tiempo. Mientras que las llamadas de señal son proporcionales al número total de personas infectadas en el distrito cuyo registro se abre en la correspondiente ventana de tiempo, aunque su confirmación de laboratorio pueda estar disponible en un momento posterior.

Para aplicar este algoritmo en PBA hemos utilizado el conjunto de datos de las llamadas telefónicas a la línea 148 COVID. Trabajamos con todas las llamadas telefónicas que entran en la línea COVID y que alcanzan el umbral de ser un contacto cercano o un caso sospechoso. La razón de la granularidad seleccionada para las alarmas es porque el distrito desde el que llama el usuario está registrado por el operador. Aunque la dirección también está registrada en principio la mayoría de las veces, en la práctica muchas ambigüedades, palabras mal escritas, u otros errores no intencionados dan como resultado que solo alrededor de ~50%-70% de los casos puede ser reconstruido correctamente. Consideramos el conjunto de datos de las llamadas entre el 1 de abril y el 26 de junio, ya que después de

esto el centro de llamadas se sobrecargó produciendo sesgos intratables. A lo largo de este período hemos ajustado los datos varias veces en diferentes conjuntos de datos, obteniendo resultados bastante similares y con un coeficiente de determinación siempre satisfactorio $R^2 > 0,85$.

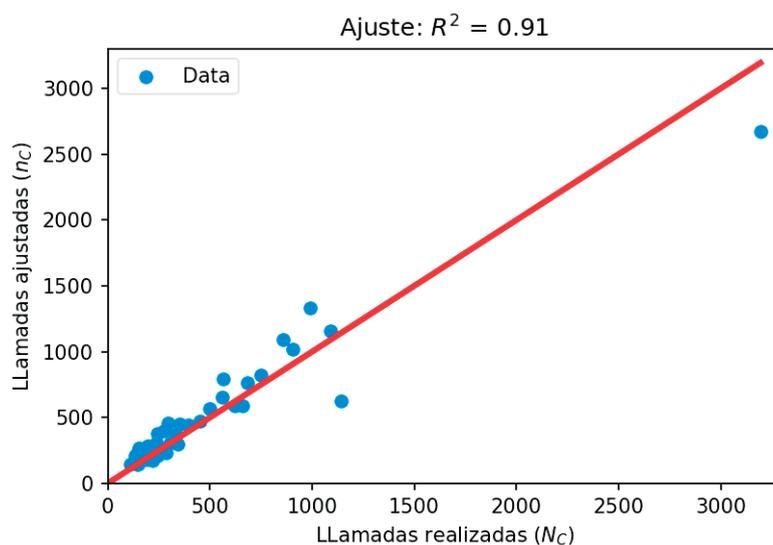
3. Resultados

3.1. Seguimiento de la epidemia a través de estimaciones de modelos

El modelo matemático descrito en la sección anterior proporciona un marco para estimar con muchos días de antelación el número de casos confirmados por laboratorio por día, en función de la distribución espacio-temporal de las llamadas telefónicas a la línea COVID. Se trata de un logro convincente porque la información de la llamada telefónica está disponible en tiempo real, mientras que la confirmación de los casos por parte del laboratorio puede requerir desde unos pocos días hasta una semana desde que los pacientes informan de sus primeros síntomas. A lo largo de esta sección mostramos cómo se puede utilizar este sistema para tener una estimación de la evolución de la epidemia en tiempo real, junto con los resultados de los casos reales en PBA.

Como este sistema fue desarrollado y aplicado en el momento, no hubo tiempo para la validación. Sin embargo, la obtención de un $R^2 > 0,85$ muy satisfactorio en el ajuste fue una señal de que el modelo funcionaba bien hasta el momento. Con el paso de los meses, tuvimos la posibilidad de comparar en una ventana de tiempo amplia la estimación del modelo con el número de casos medidos y confirmados por el laboratorio por distrito.

Gráfico 1. Llamadas realizadas versus ajustadas durante el período analizado (1 de mayo al 26 de junio dividido en dos ventanas de tiempo).



El número de llamadas total proviene del número de casos de COVID confirmados por el laboratorio usando el ajuste en la Ec. 1. El punto de datos superior derecho corresponde al distrito de La Matanza, cuya población de 1,7 millones de habitantes es al menos tres veces mayor que la del resto de los distritos.

En el Gráfico 2 mostramos la comparación entre la estimación y los casos confirmados por el laboratorio por día para dos distritos cualquiera en PBA. Se obtienen resultados similares para otros distritos. Es central observar en esta figura que el número de casos confirmados por el laboratorio (línea roja) es una información que está disponible muchos días después de la fecha correspondiente, mientras que la estimación del modelo (azul) está disponible al final de cada día. Como se puede observar en la figura, la estimación tiene un buen acuerdo con los datos reales. Hay algunos rangos de fechas en los que el hisopado de puerta en puerta a través de los agentes que realizaron vigilancia activa mediante la implementación del programa DETECTAR (Argentina.gob.ar/coronavirus/detectar) induce una subestimación esperada en los casos.

Esta vigilancia sindrómica se ha utilizado para seguir el tamaño, la propagación y el ritmo de los brotes, para vigilar las tendencias de la enfermedad y para dar garantías de que no se ha producido un posible brote. En particular, también ha sido muy útil como detección temprana de brotes, como detallamos en la siguiente sección. Los sistemas de vigilancia de los síndromes procuran utilizar los datos sanitarios existentes en tiempo real para proporcionar un análisis y una retroinformación inmediatos a los encargados de la investigación y el seguimiento de los posibles brotes. En particular, los datos recopilados por las llamadas de la línea COVID demostraron ser un aporte valioso y confiable.

El seguimiento de la epidemia a través de este modelo es especialmente útil cuando la sobrecarga de capacidad de los centros de diagnóstico provoca retrasos en la obtención de resultados. Por esta razón, tener una estimación de los casos en tiempo real y relativamente imparcial da a las autoridades de salud pública la posibilidad de tomar medidas a tiempo (Pavlin, 2013). Además, la priorización en un escenario de desastre tiene un papel principal cuando los recursos y el tiempo son limitados, como ocurrió a finales de junio en PBA. La vigilancia sindrómica basada en llamadas permitió una rápida caracterización de los diferentes distritos del PBA en cuanto a su estado epidemiológico y se tomaron las medidas consiguientes para mitigar los efectos de la epidemia.

Gráfico 2. Actividad de las alarmas en los municipios de La Matanza y Lomas de Zamora.

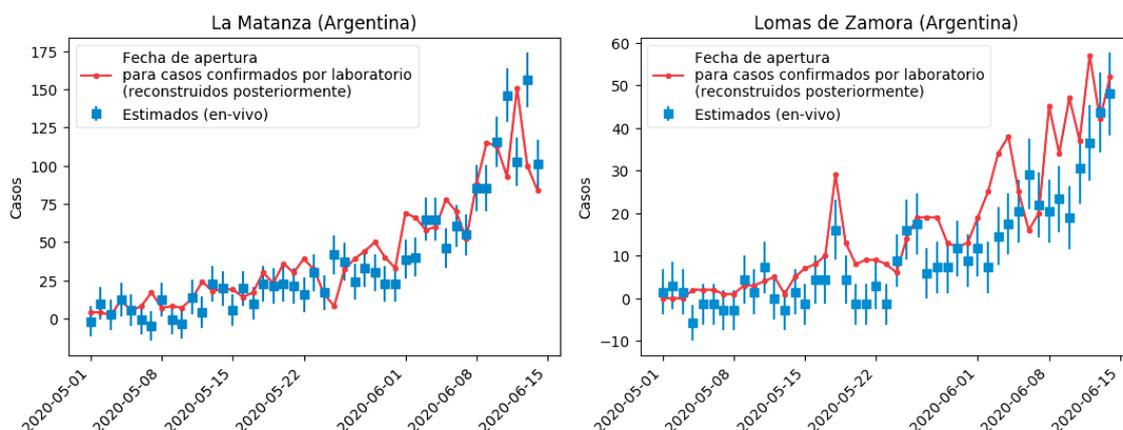


Gráfico 2: comparación de datos reales (rojo) versus estimación del modelo con una barra de error que representa un IC del 95% (azul) para dos distritos ejemplo de la provincia de Buenos Aires (La Matanza y Lomas de Zamora). La línea roja correspondiente a los casos reales confirmados con su apertura de registro en la fecha correspondiente se reconstruye muchos días después. En las fechas en las que la línea roja supera la estimación suele ser debido a que los operativos de DETECTAR (pruebas puerta a puerta) fueron llevados a cabo. En general, el modelo arroja una muy buena estimación para monitorear la epidemia en todos los distritos afectados de PBA.

3.2. Alarma de brotes tempranos

A lo largo de esta sección detallamos un convincente subproducto del modelo de la sección 2 para detectar brotes de COVID-19 considerablemente antes de producirse la confirmación del laboratorio. Describimos brevemente el funcionamiento del modelo y luego proporcionamos su detalle a través de la descripción de un caso real que ocurrió a mediados de mayo en Villa Azul y Villa Itatí en PBA.

3.2.1. Identificación de la formación de un brote

Para detectar un posible brote hay muchos indicadores que deben ser analizados simultáneamente. Por un lado, es importante tener una estimación del número absoluto y relativo diario de casos y, por otro lado, también es importante tener una estimación de la variación diaria de estos observables. Para tener un indicador cuantitativo objetivo del potencial de un brote en una región determinada, es esencial tener una evaluación correcta de las incertidumbres en todas las estimaciones del modelo. A lo largo del sistema implementado como alarma de brotes tempranos, hemos considerado la detección de anomalías una anomalía como un indicador central. Aquí la significación se define como la distancia a cero del valor central del indicador, medida en unidades de su incertidumbre. O, en otras palabras,

$$\text{significancia} = \frac{\text{valor central}}{\text{incerteza}} \quad (3)$$

El algoritmo desarrollado calcula diariamente la estimación del número total de casos nuevos en cada distrito en PBA. Dado que en la franja temporal estudiada, especialmente antes de junio, el número de casos estimados por día de muchos distritos estaba por debajo de $\sim 5 - 10$, consideramos incluir la estimación de los casos de los últimos dos días. Esto reduciría la relativa incertidumbre de Poisson debido a las pequeñas cifras. Calculamos que el número de casos estimados en valor absoluto, y también en relación con 100 mil habitantes, era igualmente sensible para todos los distritos.

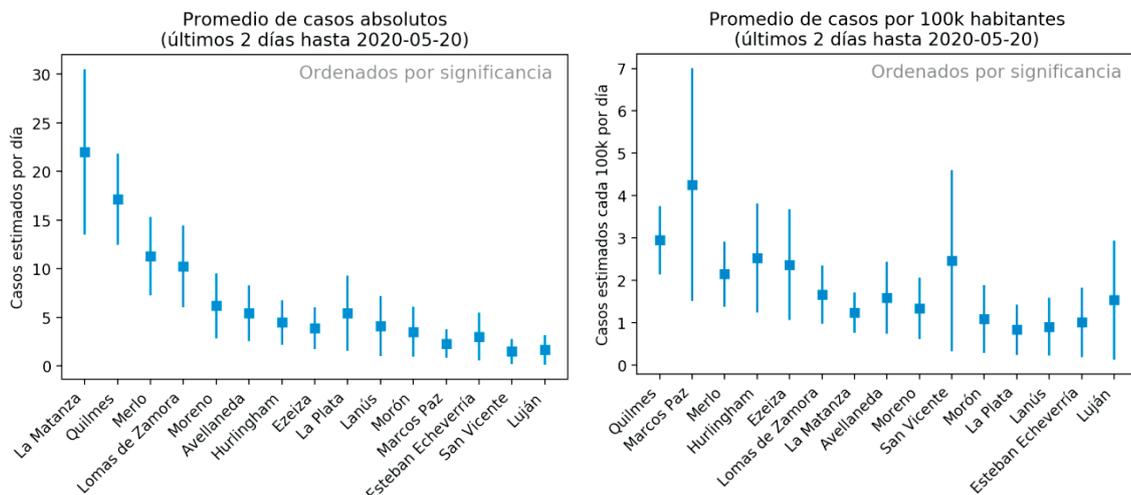
Esta alarma de brotes tempranos ha proporcionado a la administración del Ministerio de Salud de PBA herramientas muy importantes para identificar posibles brotes durante el aumento de la curva epidémica. Dado que la granularidad del algoritmo es muy pobre (distritos), el sistema debe complementarse con otros indicadores independientes, en particular los que pueden ayudar a proporcionar una localización más precisa del brote. Esto se hacía normalmente llamando manualmente a los casos registrados, y luego enviando a agentes de DETECTAR para verificar si de hecho las condiciones *in situ* serían las previstas. La alarma de brotes tempranos ha indicado muchos brotes que han sido controlados desde mediados de abril a mediados de junio. En particular, describimos en los siguientes párrafos el muy especial (ha sido cubierto en los titulares de las noticias durante varias semanas)¹ caso de Villa Azul (Quilmes) y proporcionamos los detalles de cómo la alarma temprana de brotes indicó el distrito de Quilmes.

3.2.2. Estudio de caso: Villa Azul, Quilmes

En esta sección informamos los detalles de uno de los brotes indicados por la alarma de brotes tempranos a mediados de mayo en el distrito de Quilmes. Este caso fue el primer gran brote en un barrio de bajos recursos en PBA y tuvo un gran impacto en las noticias, no solo por su magnitud sino también por su detección temprana que llevó a un estricto bloqueo y aislamiento del brote para controlar su propagación a los barrios cercanos.

¹ Se puede acceder a las noticias en el portal <https://www.argentina.gob.ar/coronavirus/detectar>

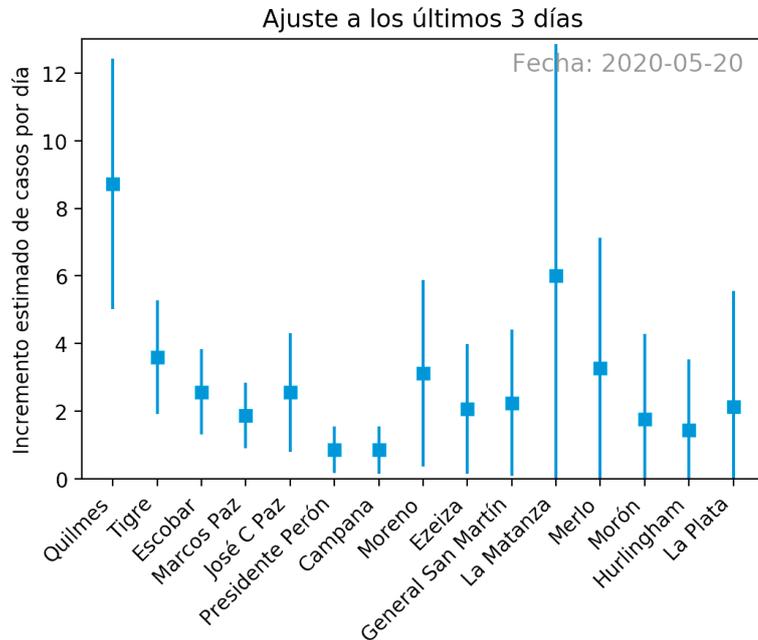
Gráfico 3. Estimación de casos por día usando el modelo de las llamadas telefónicas de la línea COVID de los últimos dos días.



Obsérvese que los distritos no están ordenados por su valor central, sino por su significado, que se define como la tasa entre el valor central y la incertidumbre. Es por eso que las barras de error son cruciales para proporcionar una alarma de brote temprano. Los resultados se muestran en valor absoluto (izquierda) y relativo a cada 100k habitantes (derecha). En el Gráfico 3 mostramos el escenario del 20 de mayo en el que Quilmes es casi tan grande como La Matanza en valor absoluto, con $\sim 1/3$ de su población. Mientras que Quilmes está en la posición superior cuando se escala a relativo por cada 100k habitantes.

El 20 de mayo la alarma indicaba un gran número de casos estimados en el distrito de Quilmes (véase el Gráfico 3), en particular Quilmes tenía la estimación más alta en número de casos por habitantes de los dos últimos días, medida a través de la significación del indicador. Además, el indicador del ajuste del aumento diario también indicaba que Quilmes era el distrito de mayor importancia (véase el Gráfico 4). Este último indicador sobre el ajuste de incremento diario a los últimos tres días puede ser visualizado en la figura 6a, donde trazamos la estimación diaria de los últimos 7 días en Quilmes. El ajuste se obtiene usando los tres últimos puntos de datos en rojo. Dado que todos estos indicadores apuntan al distrito de Quilmes, el equipo de vigilancia se encargó de localizar las llamadas telefónicas y observó un exceso procedente de Villa Azul, un barrio de bajos ingresos de Quilmes y próximo al distrito de Avellaneda.

Gráfico 4. Pendiente de un ajuste lineal a la estimación de casos por día de los últimos tres días.



La barra de error corresponde a la inclusión de la incertidumbre en cada estimación por día y en la determinación de la pendiente en el ajuste. También en esta gráfica es crucial para ordenar los distritos según la significación en esta variable. En el Gráfico 5 que vemos el 20 de mayo Quilmes en la posición superior, indica una potencial alarma temprana de un brote, como fue consecuentemente confirmado por otros indicadores unos días después.

Estas observaciones indicadas de antemano por la alarma de brotes tempranos debían ser verificadas por un indicador complementario independiente. Al día siguiente, un operativo DETECTAR fue enviado a Villa Azul, donde se verificó la situación aguda y se inició de inmediato el rastreo puerta a puerta con resultados de laboratorio urgentes. Como los primeros resultados estaban confirmando el brote en Villa Azul, la Administración del PBA decidió un estricto encierro y aislamiento durante 14 días desde el 24 de mayo.

Gráfico 5. Actividad de la alarma en el municipio de Quilmes.

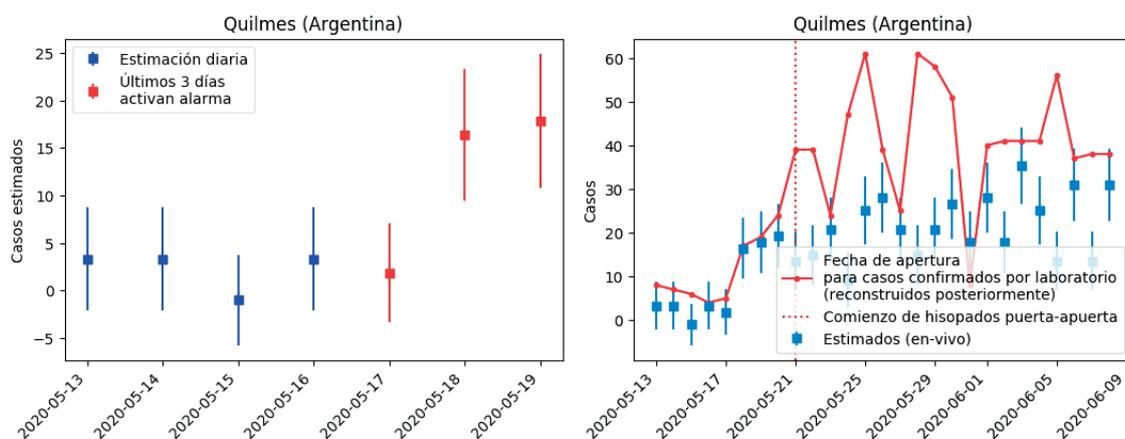


Gráfico 5, izquierda: visualización de la alarma de brote temprano el 20 de mayo para Quilmes, según lo dispuesto por el equipo de Atención de la Salud en la provincia de Buenos Aires. A la derecha: la imagen posterior incluye los casos confirmados por el laboratorio (rojo sólido), el día en que el Sistema de Salud aterrizó en Villa Azul para comenzar a probar puerta por puerta, y un rango más amplio de fechas para captar el panorama general del caso. El estricto encierro en Villa Azul sin permiso de entrada ni salida duró desde el 24 de mayo hasta el 8 de junio. Como se puede ver en la trama, durante la prueba puerta por puerta, la línea roja sólida va por encima y no se correlaciona con los casos estimados por llamadas telefónicas, como se esperaba.

3.2.3. Descripción epidemiológica y operativa de Villa Azul

Villa Azul (Quilmes) y Villa Itatí (Avellaneda) son dos barrios adyacentes de bajos recursos. El último análisis demográfico indica que Villa Azul tiene una población de 3.128 y Villa Itatí 15.142. La alta densidad de edificación y vivienda, y las pequeñas calles ponen a la población en estrecho contacto. Estas características hacen que estos barrios sean susceptibles de una rápida difusión.² Teniendo esto en cuenta, la detección temprana de los brotes implica un desafío principal en estos casos complejos en los que la detección y el bloqueo de la propagación deben hacerse cuando se reportan los primeros casos. En particular, la alarma temprana descrita anteriormente para el brote ocurrió en Villa Azul permitiendo una rápida respuesta del equipo del Sistema de Salud para mitigar y controlar su propagación a Villa Itatí.

Una vez producido el estricto cierre y aislamiento, el equipo de atención social entregó agua y alimentos. No se permitió a las personas salir de la casa durante todo el aislamiento. Los equipos sanitarios de vigilancia activa comenzaron con un monitoreo de los síntomas puerta a puerta. Se analizaron los

² Esto puede observarse en los siguientes sitios: <https://www.lanacion.com.ar/sociedad/en-fotos-villa-azul-barrio-del-sur-nid2369493/>, https://www.clarin.com/sociedad/coronavirus-buenos-aires-bloqueo-villa-azul-freno-contagios-crece-demanda-asistencia-vecinos_0_3LmgzVz0j.html, y <https://www.batimes.com.ar/news/argentina/villa-in-buenos-aires-province-isolated-due-to-covid-19.phtmlE>

casos con manifestaciones clínicas relacionadas con COVID-19. Los casos confirmados fueron aislados dentro de sus casas en los casos en que esto era posible (si había una habitación vacía, por ejemplo) y en los casos en que no lo era, las personas fueron enviadas a un centro fuera del hospital.

4. Discusión

El desarrollo del modelo matemático para estimar el número de casos de COVID-19 se hizo con urgencia y adaptándolo a los datos disponibles. No hubo tiempo de solicitar cambios en la adquisición ni en el procesamiento de los datos. Por supuesto, el algoritmo y el sistema pueden ser mejorados en muchas direcciones. En los siguientes párrafos se analizan algunas de estas características.

Una de las mayores debilidades del algoritmo es la gran granularidad, que corresponde a los distritos. Las poblaciones de los distritos en el área relevante son en promedio 500 mil personas. Esta cuestión se traduce en que la alarma de brotes tempranos deja de funcionar una vez que la densidad de casos es tal que hay más de unos pocos brotes en cada distrito. Esto ocurrió a finales de junio en PBA. En una futura implementación estamos llevando a cabo una solución para este problema obteniendo una dirección de confianza del operador entrenado por COVID que toma la llamada. Una solución más estable sería obtener esta información de la compañía telefónica; sin embargo, las regulaciones muchas veces bloquean esta posibilidad.

Por otro lado, el algoritmo tiene un beneficio muy importante que es su imparcialidad. Dado que la línea COVID funciona 24 horas los 7 días de la semana y con una metodología bastante equitativa todo el tiempo, la estimación del algoritmo no depende de la disponibilidad de pruebas o de instalaciones de prueba sobrecargadas, entre otras. Por supuesto, el sistema tiene ligeros sesgos que pueden provenir –por ejemplo– de diferentes señales de fondo debido a las diferentes características de los distritos, o al comportamiento social estacional a medida que pasan los meses. Algunos de estos sesgos pueden resolverse reajustando el modelo de vez en cuando, otros ajustando diferentes modelos en diferentes regiones.

El punto crucial del modelo matemático es que reconoce las anomalías debidas a comportamientos colectivos. Por lo tanto, encontramos que el modelo matemático y los algoritmos de Alarma de Brote Temprano pueden ser útiles para muchas otras enfermedades epidemiológicas –como, por ejemplo, el Dengue–, y otros eventos como las catástrofes naturales, entre otros. Actualmente estamos trabajando en la mejora de este sistema en muchos aspectos, incluyendo también los algoritmos de aprendizaje automático, y estos avances serán publicados en un trabajo futuro.

5. Conclusiones

Hemos creado un algoritmo de vigilancia sindrómica basado en la correlación entre las llamadas telefónicas a una línea COVID, la población de los distritos y los casos reportados. Este algoritmo funciona entendiendo que las llamadas telefónicas a una línea COVID son una parte debido a que las

personas no infectadas tienen síntomas similares (fondo) y otra parte debido a las personas infectadas (señal). Observando que el fondo tiene que ser proporcional a la población del distrito, mientras que la señal es proporcional a los casos reportados, hemos ajustado nuestra suposición. El coeficiente de determinación para la provincia de Buenos Aires (PBA) es siempre $R^2 > 0,85$ para las diferentes muestras, lo que indica la solidez de nuestra hipótesis. Además, hemos validado nuestro modelo con datos reales.

A lo largo del manuscrito se describió el modelo, sus estimaciones y cómo calculamos sus barras de error. También se ha mostrado cómo las estimaciones, que se obtienen en tiempo real, pueden ser utilizadas para abordar las políticas de salud pública sin necesidad de esperar los resultados de laboratorio, que requieren muchos más días para converger. El algoritmo funcionó en PBA de abril a junio, ya que durante este tiempo el centro de llamadas entrenado por COVID no se sobrecargó. Por lo tanto, la estimación fue relativamente imparcial.

Conflictos de interés

Ninguno

Referencias bibliográficas

- Álvarez, E. y Marsico F. (2020). COVID-19 mild cases determination from correlating COVID –line calls to reported cases–. *MedRxiv*. Recuperado de doi.org/10.1101/2020.07.26.20162008
- Ahmed, O. B. (2020). Mitigating the spread of COVID-19 in low-income countries. *Scientific Research and Essays*, 15(2), 41-44.
- Chen, D.; Cunningham, J.; Moore, K. y Tian, J. (2011). Spatial and temporal aberration detection methods for disease outbreaks in syndromic surveillance systems. *Annals of GIS*, 17(4), 211-220.
- Diwan, V.; Agnihotri, D. y Hulth, A. (2015). Collecting syndromic surveillance data by mobile phone in rural India: implementation and feasibility. *Global health action*, 8(1), 26608.
- Gao, M.; Yang, L.; Chen, X.; Deng, Y.; Yang, S.; Xu, H. y Gao, X. (2020). A study on infectivity of asymptomatic SARS-CoV-2 carriers. *Respiratory Medicine*, 169, 106026.
- Goniewicz, K. y Burkle, F. M. (2019). Disaster early warning systems: the potential role and limitations of emerging text and data messaging mitigation capabilities. *Disaster medicine and public health preparedness*, 13(4), 709-712.
- Ginsberg, J.; Mohebbi, M. H.; Patel, R. S.; Brammer, L.; Smolinski, M. S. y Brilliant, L. (2009). Detecting influenza epidemics using search engine query data. *Nature*, 457(7232), 1012-1014.
- He, X.; Lau, E. H.; Wu, P.; Deng, X.; Wang, J.; Hao, X. y Mo, X. (2020). Temporal dynamics in viral shedding and transmissibility of COVID-19. *Nature medicine*, 26(5), 672-675.
- Hope, K.; Durrheim, D. N.; d'Espaignet, E. T. y Dalton, C. (2006). Syndromic surveillance: is it a useful tool for local outbreak detection? *Journal of Epidemiology & Community Health*, 60(5), 374-375.

- Katz, R.; May, L.; Baker, J. y Test, E. (2011). Redefining syndromic surveillance. *Journal of epidemiology and global health*, 1(1), 21-31.
- Kyriacos, U.; Jelsma, J.; James, M. y Jordan, S. (2014). Monitoring vital signs: development of a modified early warning scoring (MEWS) system for general wards in a developing country. *PloS one*, 9(1), e87073.
- Krause, G.; Altmann, D.; Faensen, D.; Porten, K.; Benzler, J.; Pfoch, T. y Claus, H. (2007). SurvNet electronic surveillance system for infectious disease outbreaks, Germany. *Emerging infectious diseases*, 13(10), 1548.
- Lee, H.; Kim, K.; Choi, K.; Hong, S.; Son, H. y Ryu, S. (2020). Incubation period of the coronavirus disease 2019 (COVID-19) in Busan, South Korea. *Journal of Infection and Chemotherapy*, 26(9), 1011-1013.
- Lombardo, J.; Burkom, H.; Elbert, E.; Magruder, S.; Lewis, S. H.; Loschen, W. y Pavlin, J. (2003). A systems overview of the electronic surveillance system for the early notification of community-based epidemics (ESSENCE II). *Journal of urban health*, 80(1), i32-i42.
- Pavlin, J. A. (2003). Investigation of disease outbreaks detected by “syndromic” surveillance systems. *Journal of Urban Health*, 80(1), 107-114.
- Stoto, M. A.; Schonlau, M. y Mariano, L. T. (2004). Syndromic surveillance: is it worth the effort? *Chance*, 17(1), 19-24.
- Texier, G.; Pellegrin, L.; Vignal, C.; Meynard, J. B.; Deparis, X. y Chaudet, H. (2017). Dealing with uncertainty when using a surveillance system. *International Journal of Medical Informatics*, 104, 65-73.
- Verhagen, L. M.; de Groot, R.; Lawrence, C. A.; Taljaard, J.; Cotton, M. F. y Rabie, H. (2020). COVID-19 response in low-and middle-income countries: Don't overlook the role of mobile phone communication. *International Journal of Infectious Diseases*, 99, 334-337.