

Inteligencia artificial y ciencia de redes

LAURA VARGAS-PARADA

Laura Vargas-Parada es doctora en Investigación Biomédica Básica por la UNAM y tiene casi dos décadas de impartir cátedra en la licenciatura de Biología en la Facultad de Ciencias de la UNAM. Se dedica a la comunicación de la ciencia habiendo colaborado en diversos medios nacionales e internacionales tanto en prensa escrita como radio y televisión. Desde 2016 dirige la oficina de comunicación del Centro de Ciencias de la Complejidad (C3).

Esta publicación fue revisada por el comité editorial de la Academia de Ciencias de Morelos.

Un equipo multidisciplinario de investigadores del Centro de Ciencias de la Complejidad (C3) y el Instituto de Física de la UNAM, en colaboración con el Department of Network and Data Science de la Central European University, en Hungría, realizó un análisis utilizando inteligencia artificial y ciencia de redes para estimar el monto de evasión del impuesto al valor agregado (IVA) por la emisión de facturas electrónicas que simulan operaciones que nunca se realizaron.

“El objetivo principal de la investigación fue desarrollar métodos para identificar evasores fiscales y también para estimar cuánto se había evadido”, explicó en entrevista vía correo electrónico Carlos Gershenson, coordinador del Programa de Inteligencia Computacional y Modelación Matemática del C3, investigador del Instituto de Investigaciones en Matemáticas Aplicadas y en Sistemas y coautor del estudio.

Los académicos estiman que la evasión por emisión de comprobantes fiscales digitales por internet o CFDI fraudulentos alcanzó, durante el periodo 2015 a 2018, un promedio de poco más de 60 mil millones de pesos anuales, de acuerdo al reporte *Evasión en IVA: Análisis de redes*, que publican en un [micrositio web](#) del Servicio de Administración Tributaria (SAT), instancia para la cual realizaron el análisis. Los investigadores encontraron también que la tendencia a la evasión va en aumento, habiendo pasado de los 40 mil millones de pesos en 2015 a 77 mil millones tres años después, lo que representa un incremento del 93%. Esta evasión, de acuerdo a los investigadores, proviene de poco más de 7,677 registros federales de contribuyentes (RFC) de posibles evasores fiscales.

“Esta investigación tiene gran mérito porque utiliza técnicas de inteligencia artificial para proponer una solución de política pública a un problema sensible para países con sistemas fiscales que presentan una debilidad estructural”, dijo Roberto Ponce-López, profesor-investigador del Instituto Tecnológico y de Estudios Superiores de Monterrey, quien no participa en este estudio. “Los autores utilizan ciencia de datos para atajar un problema endémico en países en vías de desarrollo. Aplicaciones tecnológicas que cuenten con la capacidad de identificar y contextualizar herramientas desarrolladas en países industrializados para atajar problemas locales requiere de sensibilidad y creatividad”.

Para Tomás Veloz, asesor sobre ciencia de datos en la Contraloría General de la República de Chile y quien no participa en este estudio, este proyecto permite “con una inversión muy pequeña, resolver un problema que impacta de forma importante en el funcionamiento del Es-

tado y en la confianza de la gente en la administración pública”.

SISTEMAS COMPLEJOS

Para el análisis los científicos utilizaron métodos desarrollados en el estudio de los sistemas complejos que permiten analizar una gran cantidad de datos. En este caso, los datos analizados fueron todos los comprobantes digitales emitidos en México entre enero de 2015 y diciembre de 2018, anonimizados y agregados por mes. Mediante algoritmos que utilizan inteligencia artificial buscaron reproducir los patrones en la actividad de empresas que ya habían sido identificadas por el SAT como Empresas que Facturan Operaciones Simuladas (EFOS) para luego, con análisis estadístico de teoría de redes, “identificar patrones nuevos”, explicó Gershenson.

De esta forma, los científicos desarrollaron una herramienta que no solo permitió detectar patrones en el comportamiento de la emisión de comprobantes fiscales, identificando contribuyentes con comportamientos similares a los evasores fiscales (y por tanto sospechosos), sino también estimar cuántos recursos han evadido. Sobre la metodología, Gerardo Iñiguez, profesor asistente en la Central European University, investigador visitante de la Aalto University, Finlandia, y coautor del estudio, explicó en un mensaje electrónico: “Utilizamos dos métodos basados en algoritmos de aprendizaje de máquinas para detectar posibles EFOS aún no identificados [por el SAT] a partir de una lista de EFOS ya detectados”.



En el primer método de aprendizaje automatizado, los investigadores utilizaron las características de EFOS ya detectados como evasores para entrenar una red neuronal. Una vez que la red neuronal está entrenada, se aplica el mismo algoritmo a todos los RFCs del estudio, encontrando los RFCs más similares a los EFOS ya detectados y, por tanto, sospechosos de la misma actividad de evasión. “El segundo método actúa de forma similar”, añade Iñiguez, especialista en ciencia social computacional. En este caso, los EFOS ya detectados se usan para entrenar un bosque aleatorio, el cual informa de las características que cualquier otro RFC debe tener para ser considerado sospechoso de ser EFOS. El aprendizaje de los dos algoritmos (redes neuronales y bosques aleatorios) resulta de la acumulación de información estadística que es lo que permite identificar a los EFOS sospechosos. “Si algún RFC se comporta estadísticamente de forma similar o presenta características similares a los EFOS ya identificados, entonces los algoritmos lo detectan como un RFC sospechoso de ser EFOS”, explica Iñiguez, físico y doctor en ciencias computacionales. Usar dos métodos distintos permite a los investigadores contrastar resultados y hacer más robusta su investigación.

LIMITACIONES

Para Iñiguez, una de las limitaciones de su investigación es que no permite estimar la evasión fiscal en actividad económica que no deja rastro en datos del SAT,



C
I
E
N
C
I
A

launion.com.mx

@uniondemorelos

ESTA PUBLICACIÓN FUE REVISADA POR EL COMITÉ EDITORIAL DE LA ACADEMIA DE CIENCIAS DE MORELOS

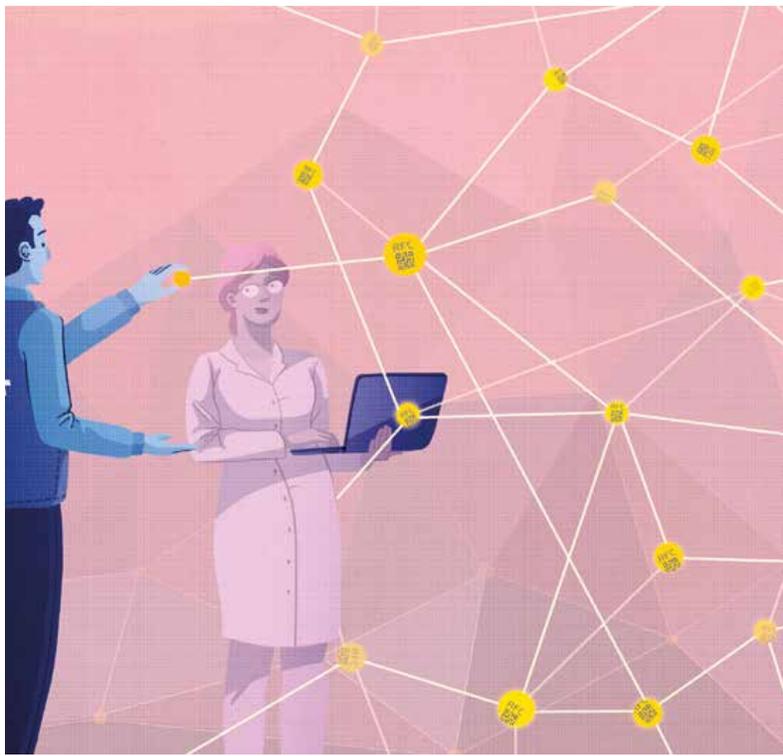
Para actividades recientes de la academia y artículos anteriores puede consultar: www.acmor.org.mx
¿Comentarios y sugerencias?, ¿Preguntas sobre temas científicos? CONTACTANOS: editorial@acmor.org.mx

- » BIBLIOGRAFÍA
- » Evasión en IVA: Análisis de redes. [en IVA Análisis de Redes.pdf](#)
- » Recupera el SAT 28 mil millones [en el SAT](#)
- » <https://www.eluniversal.com.mx>



ACADEMIA DE CIENCIAS DE MORELOS, A.C.

contra la evasión fiscal



El análisis estadístico de teoría de redes implica definir una red donde cada nodo es un RFC y un enlace dirigido describe un comprobante fiscal (CFDI) entre dos contribuyentes. La estructura de la red indica entonces toda la actividad económica entre RFCs en México de acuerdo al SAT, mes por mes en varios años. Al estudiar las propiedades estadísticas de la estructura de esta red (medidas de conectividad, centralidad, etc.) es posible identificar el patrón típico de nodos y enlaces que se encuentra alrededor de un EFOS ya identificado por el SAT. Buscando otros RFCs en la red que tengan una estructura similar a su alrededor es lo que permite identificar los RFCs sospechosos de ser EFOS, explicó Iñiguez.

Por lo anterior, los investigadores consideran que en un futuro y en colaboración con el SAT, será necesario realizar un análisis más avanzado de teoría de redes con el fin de detectar "patrones típicos de colaboración entre contribuyentes (motifs temporales) que estén relacionados con actividades de evasión". Un motif es un patrón específico entre un número pequeño de nodos (por ejemplo, 3 conexiones entre 3 nodos: un triángulo). Un motif temporal es un motif que también tiene información de los momentos en que los enlaces (en este caso transacciones entre RFCs) aparecen.

La idea es buscar estos patrones específicos que aparecen comúnmente alrededor de un EFOS (por ejemplo, un triángulo o algo más complicado) y buscar otros RFCs que tengan patrones similares en la red a su alrededor. De esta forma puede medirse de forma específica cómo los EFOS interactúan con otros RFCs, buscar esa misma estructura en otras partes de la red, y así identificar con más detalle las prácticas típicas que promueven la evasión fiscal, como el uso de RFCs genéricos y la auto-facturación en círculos de evasores. Esto será fundamental para "crear sistemas automáticos de detección y monitoreo de EFOS sospechosos", explica Iñiguez. Un aspecto importante de esta investigación es que una vez que se entrena al sistema también es posible identificar nuevos patrones de evasión. "Nuestro análisis nos permitirá identificar (en el futuro) métodos de evasión más sofisticados. Uno de ellos es circularidad, digamos que el siguiente paso en complejidad de la auto-facturación: en vez de facturarse a sí mismo, un EFOS podría facturar a otro RFC, que factura a otro RFC, y así varias veces hasta facturar al EFOS inicial (formando un círculo), lo cuál hace la identificación de evasión fiscal más difícil. Este tipo de comportamientos o patrones estadísticos (de hecho un motif) es algo que puede ser descubierto a través de herramientas de análisis de redes como las que nosotros estamos usando", explicó Iñiguez.

RELEVANCIA

La relevancia de este estudio es "gigante", dice Veloz. "Desde un punto de vista histórico, la poca integración de la ciencia en el quehacer del mundo público ha sido un obstáculo importante para el desarrollo de los países y del acercamiento a la justicia social", escribió en mensaje electrónico el también director de la Fundación para el Desarrollo Interdisciplinario de la Ciencia, la Tecnología y las Artes, en Santiago de Chile. "Es urgente vincular la ciencia y el mundo académico en general en los procesos del mundo público para poder lograr resultados más eficientes."

La ciencia de los datos me parece que es la llave para permitir dicha vinculación, pues usando las herramientas de esta investigación se puede agilizar tremendamente el trabajo, e incluso, lograr resultados que ningún equipo de personas podría, pues el procesamiento de datos masivos está fuera del alcance de la información que somos capaces de manejar como seres humanos".

Para el especialista en modelamiento matemático, ciencia de datos y estudios interdisciplinarios, este tipo de análisis es de los más "comprobadamente útiles": el uso de datos masivos para clasificación automatizada de conductas ilegales. Para el académico chileno, con la información que provee esta herramienta, el personal del SAT podrá ganar eficiencia en su trabajo. Algo conocido como inteligencia de negocio. "Con una inversión de recursos relativamente pequeña, los organismos públicos pueden vincularse a científicos de datos que permitan realizar inteligencia de negocios y mejorar tremendamente sus procesos. En el futuro esperamos que esto se establezca para poner el conocimiento especializado al servicio de los procesos del Estado", dijo.

CIENCIA PARA RESOLVER PROBLEMAS

A principios de junio, las autoridades fiscales dieron a conocer que habían logrado recuperar casi 28 mil millones de pesos por evasión y fraude, de acuerdo a una nota publicada por el periódico El Universal. Las empresas identificadas por el SAT, para evitar ser procesadas, debieron reparar el daño ocasionado al erario público además de aceptar públicamente su mal proceder a través de edictos publicados en periódicos de circulación nacional. La identificación de RFCs sospechosos permitió al SAT dar un paso en contra de la evasión fiscal. Y aunque este órgano administrativo no lo ha dicho públicamente, es muy probable que parte de estos logros se deban a la acertada decisión que tomó de invertir en investigación. Un ejemplo para otras instancias en el país que podrían beneficiarse del conocimiento científico y tecnológico para la resolución de problemas de interés nacional.

Esta columna se prepara y edita semana con semana, en conjunto con investigadores morelenses convencidos del valor del conocimiento científico para el desarrollo social y económico de Morelos. Desde la Academia de Ciencias de Morelos externamos nuestra preocupación por el vacío que genera la extinción de la Secretaría de Innovación, Ciencia y Tecnología dentro del ecosistema de innovación estatal que se debilita sin la participación del Gobierno del Estado.

Nota: Una versión previa de este artículo se publicó en la sección de Noticias del Centro de Ciencias de la Complejidad (C3) de la UNAM. <https://www.c3.unam.mx/noticias/noticia174.html>

como es el caso de la economía informal. En México, el 56% de los empleos son informales y generan el 22% del Producto Interno Bruto (PIB) explicó Ponce-López, especialista en combinar herramientas de análisis espacial con machine learning y bases de datos de gran escala para construir infraestructura informacional, vía correo electrónico. El problema con la evasión fiscal y la informalidad, dijo, es que limitan los recursos que el Estado podría obtener para incrementar las inversiones en salud, educación, infraestructura y programas sociales.

Otra limitación muy relevante, aclara Iñiguez, es que "existe la posibilidad de que algunos contribuyentes honestos se comporten estadísticamente de forma similar a EFOS ya identificados, y por tanto sean clasificados erróneamente como EFOS sospechosos". Por ello, para Gershenson, especialista en sistemas evolutivos y adaptativos nos dice que "este tipo de herramientas no reemplazan a los expertos del SAT pero les pueden ayudar a identificar más evasores en menos tiempo, con lo que pueden potencialmente recuperar más recursos públicos", haciendo más eficiente su lucha contra la evasión fiscal.

Ponce-López considera que una limitante que enfrenta este tipo de estudios es "obtener más registros de los llamados "factureros" para incrementar el tamaño de la base de entrenamiento de la herramienta". El académico considera que esto podrá solventarse conforme mejore la disponibilidad de información digital robusta y de calidad.

Veloz, quien es también director de modelamiento matemático del Center Leo Apostel for Interdisciplinary Studies en la Vrije Universiteit, en Bélgica, considera que la aplicación de aprendizaje automático en clasificación tiene algunas limitaciones, como la dependencia

del método en los datos, esto porque los datos pueden tener múltiples sesgos que ocasionen que la inteligencia artificial aprenda de forma sesgada lo que es una operación simulada. Esto ocasionaría que el sistema podría no detectar ciertos tipos de operaciones simuladas o, como lo mencionó Iñiguez, podría clasificar como sospechosas ciertas operaciones que no lo son. "Para esto se debe trabajar continuamente en mejorar la manera en que la inteligencia artificial aprende, y eso, puede tomar bastante trabajo", explica el académico. Una limitante más, en opinión de Ponce-López, es la capacidad institucional para poder aprovechar este tipo de herramientas, ya que sólo podrá ser útil "en la medida en que el SAT haga uso de ella", lo cual requiere importantes cambios organizacionales y de procedimientos. "El cambio organizacional suele ser un reto aún más complicado que el reto tecnológico", agrega el académico en entrevista vía correo electrónico.

Ponce-León destaca, sin embargo, que la investigación es en sí misma ya una muestra del grado de colaboración logrado entre los autores y el personal encargado de recolectar y clasificar la información fiscal de millones de contribuyentes. "El análisis y la especificación de los distintos modelos ilustra un grado importante de colaboración entre la parte técnica y los funcionarios públicos. Esto último representa un caso exitoso de colaboración y transformación organizacional", menciona el experto.

TEORÍA DE REDES PARA IDENTIFICAR OTROS PATRONES DE EVASIÓN

http://omawww.sat.gob.mx/gobmxtransparencia/Paginas/documentos/estudio_opiniones/Evasion

de pesos de empresas evasoras.
[x/cartera/recupera-el-sat-28-mil-millones-de-pesos-de-empleados-evasoras](http://cartera/recupera-el-sat-28-mil-millones-de-pesos-de-empleados-evasoras)

