

## Análisis de accidentes de tráfico basado en Spark e inferencia causal

Quanjin Liu

Facultad de Ciencias de la Computación, Universidad de Wuhan

2019302110266@whu.edu.cn

**Resumen.** En los últimos años, los accidentes de tráfico han ocurrido con frecuencia, causando grandes pérdidas para la seguridad personal y patrimonial. El estudio de los datos sobre accidentes de tráfico ayuda a identificar los factores clave a partir del big data. Este artículo procesa y calcula el big data con base en la plataforma Spark. Mediante la introducción de la inferencia causal en el análisis de accidentes de tráfico, se establecen las relaciones causales entre 17 factores y la gravedad de los accidentes, analizando así las causas raíz e intermedias de los accidentes. Además, se realiza un estudio de intervención para evaluar la influencia de cada factor. El estudio concluye que las condiciones físicas de los peatones y las condiciones meteorológicas son las causas raíz.

Las demás son causas intermedias. Además, la presencia policial y la reducción del volumen de tráfico se consideran las mejores maneras de reducir los accidentes de tráfico. Por lo tanto, este artículo considera que, en la práctica, deberíamos reducir la incidencia de accidentes de tráfico controlando el flujo vehicular y aumentando las patrullas o la fuerza policial de servicio. Estos hallazgos proporcionan una base científica para que los departamentos de gestión del tráfico desarrollen estrategias de seguridad vial más eficaces.

**Palabras clave:** accidente de tráfico, big data, spark, inferencia causal, estudio de intervención

### 1. Introducción

#### 1.1 Antecedentes de la investigación

Durante el último siglo, las ciudades han experimentado un desarrollo continuo. Como resultado, las carreteras se han vuelto cada vez más intersectoriales y los automóviles se han convertido gradualmente en una parte esencial de la vida cotidiana de los ciudadanos. Actualmente, se ha establecido un sistema de transporte complejo, con vehículos privados por doquier. Sin embargo, este fenómeno provoca congestiones de tráfico y accidentes frecuentes, lo que a su vez resulta en enormes pérdidas personales y materiales. Las pérdidas económicas causadas por accidentes de tráfico ascienden a unos 30 000 millones de dólares anuales, lo que equivale aproximadamente al 3 % del producto interior bruto (PIB) de los países de todo el mundo[1]. Por lo tanto, cómo reducir eficazmente la incidencia de accidentes de tráfico, cómo responder con rapidez tras un accidente y cómo reducir las pérdidas derivadas de un accidente se han convertido en problemas importantes.

#### 1.2. revisión de la literatura

En los últimos años, los investigadores han utilizado ampliamente el big data para analizar los datos de accidentes de tráfico y encontrar causas y sugerencias. Esta tecnología puede proporcionar nuevas soluciones de forma eficiente para explorar las causas de los accidentes. Además, puede ayudar a enriquecer los métodos de análisis de accidentes existentes y a evitar la necesidad de análisis.

Sesgo debido a la experiencia subjetiva de los analistas[2]. Muchos académicos han analizado con éxito las diversas causas de accidentes utilizando big data y tienen cierto grado de confianza en la predicción de su ocurrencia. Luo Yulin construyó un modelo de diagnóstico de accidentes combinando bosque aleatorio, árbol de decisión y red neuronal de regresión generalizada, estimó cuantitativamente la influencia de cada factor y generó la clasificación de gravedad del accidente. Finalmente, este último tuvo una mayor tasa de éxito de predicción y se apoyó en ejemplos[2]. Zheng Lai et al. establecieron un árbol de fallas difuso TS con accidentes de tráfico graves como evento principal, personas, vehículos, carreteras y medio ambiente como eventos intermedios, y 24 subfactores como eventos básicos. Lo convirtieron en una red bayesiana y luego infirieron bidireccionalmente la importancia y la probabilidad posterior de los eventos básicos para determinar las causas principales. La precisión y la fiabilidad de los resultados del análisis de causas de accidentes de tráfico graves se mejoraron mediante razonamiento directo e inverso[3]. Han Tianyuan et al. construyeron un modelo jerárquico del mecanismo de accidentes de tráfico graves e importantes basado en minería de texto. Los resultados muestran que la contribución de la red de causas de accidentes de tráfico graves a la red de causas de accidentes de tráfico graves se compone de conductas ilegales, riesgos para la seguridad y operaciones inadecuadas, de mayor a menor. La combinación de las causas directas de conductas ilegales y operaciones inadecuadas con las causas indirectas de riesgos para la seguridad es la razón fundamental de la inestabilidad del sistema de operaciones de seguridad de accidentes graves[4].

En cuanto a la tecnología de big data, Spark es especialmente adecuado para el sector del transporte gracias a sus características de computación basada en memoria. Spark agiliza el procesamiento de datos distribuidos, que son precisamente el tipo de datos de accidentes de tráfico. Muchos investigadores han utilizado Spark para realizar investigaciones en el sector del transporte. Ebtesam Alomari et al. propusieron un método para detectar automáticamente eventos relacionados con el tráfico vial a partir de tuits en dialecto saudí mediante aprendizaje automático y big data[5]. Saraswathi et al. implementaron un sistema de monitoreo de tráfico en tiempo real basado en Spark. En el artículo, se predice el volumen de tráfico mediante vehículos conectados, se procesan datos de transmisión en tiempo real con Apache Spark y se muestra en un panel de control con Springboot[6]. Guo Yuda et al. diseñaron e implementaron un algoritmo paralelo eficiente basado en el marco de computación Spark para la segmentación de la red vial y el cálculo de la densidad kernel en la estimación de la densidad kernel de la red vial.

Tomando como ejemplo los accidentes de tráfico, se realizaron cuatro grupos de experimentos para realizar un análisis comparativo. Los resultados muestran que el algoritmo paralelo de estimación de densidad de kernel de la red vial, basado en el marco de computación Spark, presenta una alta eficiencia computacional y buena escalabilidad[7]. Spark puede facilitar el procesamiento de una gran cantidad de datos sobre accidentes de tráfico en este experimento. Este artículo intenta estudiar los accidentes de tráfico mediante inferencia causal en la plataforma Spark.

### 1.3. Brechas de investigación

Es cierto que las investigaciones existentes han analizado el impacto de diversos factores en los accidentes de tráfico y permiten obtener la ponderación de cada uno. Sin embargo, la mayoría de los estudios previos emplearon métodos tradicionales, como redes neuronales y análisis de regresión, que carecían de un análisis profundo de las relaciones causales, lo que dificultaba la identificación de las verdaderas causas de los accidentes entre los diversos factores. Mientras tanto, muchos estudios se centraron en la relación entre un solo factor y los accidentes de tráfico, pero la dimensión de los datos es limitada e insuficiente para reflejar la situación general. Además, la mayoría o una gran parte de la investigación actual utilizó muestras de datos limitadas, que pueden no representar con precisión la situación general. La investigación existente no ha explorado el potencial de mejora mediante intervenciones específicas, lo cual constituye una brecha significativa en la literatura actual. Si se introduce la inferencia causal en la investigación sobre datos de accidentes de tráfico y se utilizan redes bayesianas (BN) para introducir intervenciones externas, se puede ayudar a definir los efectos causales de las intervenciones externas y describir las relaciones causales entre múltiples variables relacionadas con los accidentes[8]. En consecuencia, ayuda a obtener conclusiones sobre cuánta reducción de accidentes se puede lograr con cierta intervención, identificando así eficazmente las prioridades de mejora en la práctica de prevención de accidentes y guiando las actividades relevantes.

La inferencia causal es una línea de vanguardia en la actualidad y aún no se ha aplicado en el análisis y la predicción de accidentes de tráfico. Sin embargo, ha tenido éxito en campos como la medicina y la educación. Zhang Yu resolvió el sesgo de autoselección en el ámbito educativo basándose en la inferencia causal[9].

Li Shiyuan et al. estudiaron la relación causal entre la participación en tutorías extracurriculares y

La generación de emociones negativas en estudiantes de secundaria en China continental se basó en dos fases de datos de la Encuesta Longitudinal de Educación de China, y se respondieron preguntas como "si la participación en tutorías extracurriculares causa depresión" y "quién está deprimido" a nivel cuantitativo[10]. Liu Xinhui et al. partieron de datos realistas y utilizaron una serie de métodos de inferencia causal para analizar indicadores de salud que presentan relaciones causales basadas en evidencia con resultados de salud/enfermedad, lo que puede proporcionar evidencia práctica y valiosa para el manejo de la salud/enfermedad[11].

#### 1.4. Tema y método de investigación

En este estudio, los pasos de la investigación se dividen en dos componentes distintos: la fase de procesamiento de datos, que implica la preparación de los datos para el análisis, y la fase de inferencia causal, que implica la identificación de las relaciones causales entre las variables.

Este experimento utiliza los datos de accidentes de tráfico de Estados Unidos de 2005 a 2007 como conjunto de datos y utiliza Pandas para procesar los datos relevantes. Este conjunto de datos tiene 33 dimensiones. Primero, se predice que se eliminarán los datos irrelevantes del vínculo de accidentes de tráfico y, a continuación, se eliminarán o completarán los valores faltantes.

Posteriormente se indica el tipo de datos del objeto a normalizar y finalmente se obtiene un conjunto de datos que puede identificar relaciones causales.

Tras el procesamiento anterior, se introduce el CasualModel de la biblioteca Dowhy para la identificación de la relación causal, y se utiliza Matplotlib para la visualización y así obtener el diagrama de relación causal de todo el conjunto de datos. Mediante estas operaciones, se puede determinar qué factor debería ser la causa, cuál el efecto y la relación entre ellos.

Tras analizar la relación causal a partir del diagrama de relación causal obtenido, este artículo continuará descubriendo los resultados de las intervenciones. Las variables se estiman en la relación causal obtenida para obtener el peso de influencia de cada factor, evaluando así los efectos de la intervención.

Finalmente, para evaluar la precisión y autenticidad del experimento, se realiza una prueba de refutación.

## 2. Experiment

### 2.1. tratamiento de datos

Este experimento utiliza un conjunto completo de datos sobre accidentes de tráfico en Estados Unidos, que abarca el período de 2005 a 2007, y que sirve de base para nuestro análisis. Este conjunto de datos tiene múltiples dimensiones, abarca un amplio rango de datos y contiene una gran cantidad de información. Los tipos de datos específicos se muestran en la tabla 1.

Tabla 1. Tipos de datos no procesados

Índice de accidentes	objeto	Ubicación_Este_OSGR	flotar
Gravedad del accidente	entero	Ubicación_Northing_OSGR	flotar
Número_de_vehículos	objeto int	¿Asistió el oficial de policía a la escena del accidente?	
Número de víctimas	entero	Autoridad local (autopista)	objeto
Condiciones de luz	objeto	Cruce de peatones - Control humano	objeto
Condiciones meteorológicas	objeto	Cruce de peatones - Instalaciones físicas	objeto
Peligros en la calzada	objeto	Condiciones de la superficie de la carretera	objeto
Área urbana o rural	entero	Condiciones especiales en el sitio	objeto
Autoridad local (distrito)	entero	LSOA_de_Ubicación_del_Accidente	objeto

Tabla 1. (continuación).

Longitud	flotar	Fuerza policial	entero
Latitud	flotar	Día de la semana	entero
Fecha	objeto	Tipo de carretera	objeto
Tiempo	objeto	Límite de velocidad	entero
Año	entero	Detalle de la unión	flotar
Control de cruces	objeto	1er número de carretera	entero
1.ª clase de carretera	entero	2do_Número_de_Carretera	entero
2da clase de carretera	entero		

De la tabla anterior, las estadísticas de la dimensión de datos se componen de 5 elementos de tipo float, 13 elementos de tipo int y 15 elementos de tipo object. Este conjunto de datos incluye factores, algunos de los cuales se consideran causas de accidentes de tráfico, como factores humanos, peatones, factores ambientales, clima, etc. Otros factores son criterios de evaluación de la gravedad de los accidentes de tráfico, como la clasificación de gravedad y el número de heridos. El conjunto de datos contiene un total de 570.000 datos de accidentes de tráfico.

Por lo tanto, el conjunto de datos puede satisfacer los requisitos experimentales y la conclusión final también tiene valor de referencia.

Hay algunas variables irrelevantes en este conjunto de datos, como el ID y el año del accidente, que no son muy útiles para estudiar la causa del mismo. Además de estos datos, los valores faltantes también constituyen una parte importante del conjunto de datos y también deben procesarse. Este artículo utiliza los métodos drop() y dropna() de la biblioteca Pandas de Python para procesar las variables irrelevantes y los valores faltantes.

Posteriormente, dado que la biblioteca Dowhy utilizada para la inferencia causal no puede procesar el tipo de objeto, se debe utilizar un método eficaz para convertir los datos de tipo de objeto del conjunto de datos a tipo entero o flotante. El método LabelEncoder de la biblioteca sklearn.preprocessing puede ser útil. Utiliza un valor entre 0 y el número de categorías - 1 para codificar la etiqueta de destino y convertir datos no numéricos en numéricos. Este artículo procesa 15 datos de tipo de objeto mediante este método. Tras el procesamiento de datos descrito anteriormente, el conjunto de datos final consta de 18 dimensiones y 315.000 datos de accidentes de tráfico. Los tipos de datos procesados se muestran en la tabla 2.

Tabla 2. Tipos de datos procesados

Día de la semana	entero	Fuerza policial	entero
Gravedad del accidente	entero	Tiempo	flotar
Número_de_vehículos	entero	¿Asistió un agente de policía a la escena del accidente?	entero
Número de víctimas	entero	Límite de velocidad	entero
Condiciones de luz	entero	Cruce de peatones - Control humano	entero
Condiciones meteorológicas	entero	Cruce de peatones - Instalaciones físicas	entero
Peligros en la calzada	entero	Condiciones de la superficie de la carretera	entero

Tabla 2. (continuación).

Área urbana o rural	entero	Condiciones especiales en el sitio	entero
Tipo de carretera	entero	Control de cruces	entero

## 2.2. Identificación de relaciones causales

En realidad, deben existir ciertas relaciones causales entre las variables que provocan los accidentes de tráfico.

Sin embargo, es difícil construir un gráfico de relación causal correcto basándose únicamente en el conocimiento previo.

La biblioteca de inferencia causal (Dowhy) puede facilitar el descubrimiento de relaciones causales. Los métodos de inferencia causal (PC) y de inferencia causal general (GES) son métodos de descubrimiento causal rápidos y ampliamente utilizados. Por lo tanto, este artículo intenta descubrir el grafo de relación causal mediante la introducción de los métodos PC y GES, y visualizarlo mediante las bibliotecas GraphUtils y pyplot.

Existe una diferencia significativa entre los diagramas causales obtenidos mediante los métodos PC y GES. El algoritmo PC fue propuesto por Peter Spirtes et al. Se basa en pruebas de independencia condicional. La idea central del algoritmo PC es construir un gráfico causal observando los pares de variables en los datos para determinar si están relacionados directa o indirectamente, basándose en pruebas de independencia condicional. Por otro lado, el algoritmo GES asume que no existe una relación causal entre la mayoría de las variables e intenta encontrar una estructura causal dispersa. Al analizar la relación entre la gravedad de los accidentes de tráfico y las condiciones de luz, GES cree que la gravedad de los accidentes de tráfico causa las condiciones de luz. Por lo tanto, al comparar los gráficos causales obtenidos por PC y GES, PC se ajusta más al conocimiento previo y se corresponde mejor con la realidad. Sin embargo, según el conocimiento previo, las condiciones de luz son condiciones objetivas, y la relación causal entre ambas debería ser que, potencialmente, las condiciones de luz causan accidentes de tráfico hasta cierto punto. Por lo tanto, existe una contradicción entre el diagrama causal GES y el conocimiento previo. Mientras que, la relación causal analizada por PC es consistente con el conocimiento previo. En este artículo se analiza la relación causal obtenida mediante el método PC.

El gráfico causal ilustra claramente dos nodos: Paso de peatones-Instalaciones físicas y Condiciones meteorológicas, como nodos iniciales, que representan las causas raíz. El cuerpo policial involucrado en la gestión del accidente es Fuerza policial, y la gravedad del accidente es Gravedad del accidente, que son los nodos finales y los resultados finales de todo el diagrama de causa-efecto.

La gravedad del accidente se ve afectada directamente por las condiciones especiales del lugar, el control de cruces, el número de vehículos, las condiciones de la luz, el límite de velocidad, el número de víctimas, el área urbana o rural, etc.

El agente de policía que acude al lugar del accidente, las condiciones meteorológicas y otras variables, y se ve afectado indirectamente por otras variables, demuestra que los accidentes de tráfico son resultado de múltiples factores.

El nodo Police\_force se ve afectado directamente por algunas variables como instalaciones físicas para cruces de peatones, control de cruces, condiciones de la superficie de la carretera, área urbana o rural, ¿actuó la policía?

Oficial de Asistencia al Lugar del Accidente, y también se ve afectado indirectamente por otras variables. Estas

Las relaciones revelan que cuando los factores que provocan accidentes de tráfico son diferentes, la fuerza policial enviada cambiará en consecuencia.

Además, el número de víctimas, una variable a la que la gente presta atención, está relacionada con las condiciones de la superficie de la carretera, el límite de velocidad, si el agente de policía asistió a la escena del accidente y el estado de la luz.

iones, Control de Cruces, Número de Vehículos, Zona Urbana o Rural. Muestra que la probabilidad de que un accidente cause víctimas está potencialmente relacionada con el estado de la calzada, el límite de velocidad, la presencia policial, las condiciones de iluminación, el correcto funcionamiento de los semáforos en la intersección, el volumen de tráfico y si se trata de una zona urbana.

## 2.3. Estimación

Si el cambio de una variable conlleva un cambio en la variable dependiente final, se puede definir que la variable provocará la ocurrencia de la variable dependiente. En este proceso, todo lo demás permanece inalterado.

Por tanto, en este paso, este artículo identifica la relación causal y el efecto a estimar a través de las propiedades del gráfico causal.

En este artículo, se estiman todos los factores, excepto la Gravedad del Accidente, como parámetros de intervención. Primero, se asignan los parámetros de tratamiento y resultado a la variable a estimar y a la variable dependiente, respectivamente, y se establece un modelo causal para la variable a estimar y la gravedad del accidente (Gravedad del Accidente). A continuación, se utilizan los métodos de identificación del efecto (`identify_effect()`) y estimación del efecto (`estimation_effect()`) del modelo causal para la estimación. Este artículo estima la intervención de cada factor mediante el criterio de puerta trasera del modelo causal estructural.

En la etapa de estimación, los valores estimados de otras variables se muestran en la figura 1

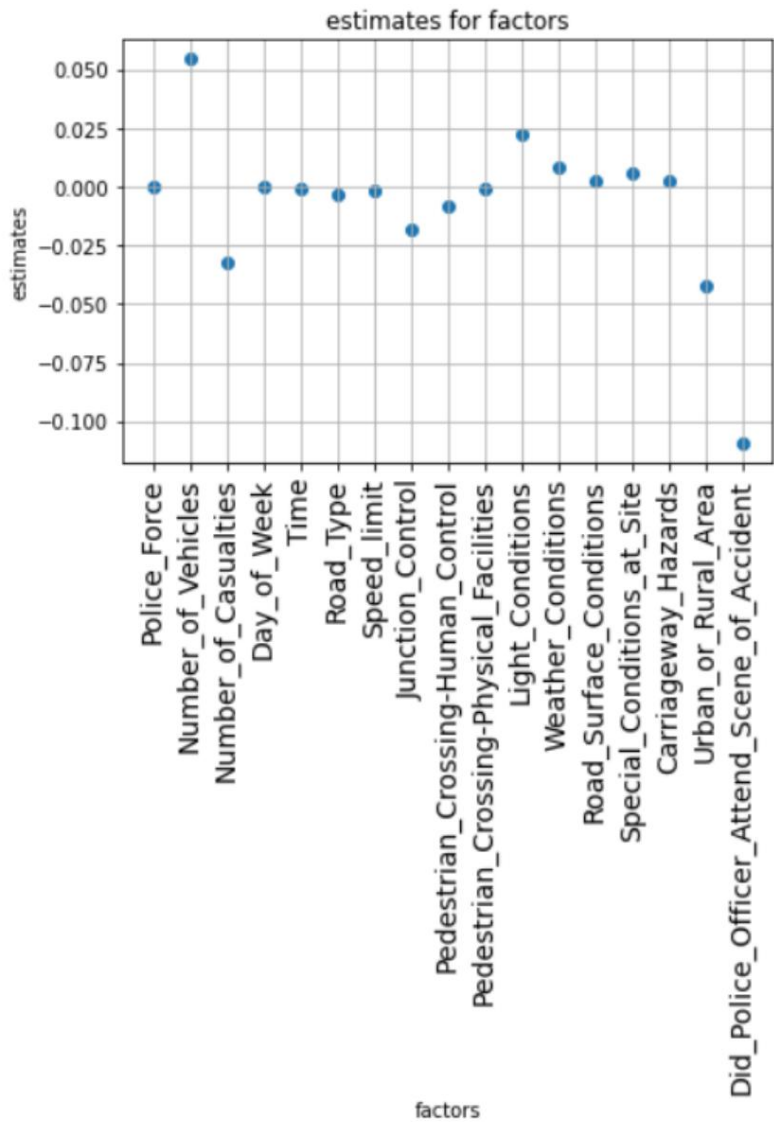


Figura 1. Valores estimados.

Número de vehículos, Número de víctimas, ¿Asistió algún agente de policía al lugar del accidente?, Urb. Las variables `an_or_Rural_Area` y `Light_Conditions` presentan estimaciones significativamente mayores de la gravedad de los accidentes que otras variables. Esto demuestra que el volumen de tráfico, el número de víctimas, la presencia policial en el lugar de los hechos, la ubicación en una zona urbana o rural, y la calidad de la iluminación...

Existe una relación causal significativa con la gravedad de los accidentes de tránsito. Es necesario que otros factores influyan actuando sobre estos cinco factores.

2.4. Prueba de refutación

Para verificar la fiabilidad de los resultados de la estimación causal, realizamos una prueba de refutación. De hecho, la relación causal en el diagrama causal obtenido en el segundo paso es solo una hipótesis sobre diversos factores de este artículo. La prueba de refutación busca verificar si esta hipótesis es correcta.

La idea básica de la prueba de refutación es utilizar el método de causa común aleatoria. Las causas comunes aleatorias son causas aleatorias e inespecíficas que prevalecen en los datos observados y afectan a múltiples variables.

Estas causas suelen ser impredecibles e interfieren con las inferencias sobre las relaciones causales. En este artículo, se utiliza este método para eliminar la influencia de estas interferencias. Su funcionamiento específico consiste en añadir covariables aleatorias a los datos y repetir el análisis para comprobar si la estimación causal cambia. Si la hipótesis inicial es correcta, la estimación causal no debería cambiar significativamente.

Este artículo utiliza herramientas estadísticas para evaluar la interferencia causada por esta variable aleatoria. Entre ellas, el valor p es un concepto clave. El valor p es un parámetro estadístico que se utiliza para evaluar la significancia de la diferencia entre los datos observados y la hipótesis nula. Representa la probabilidad de observar la muestra actual o casos más extremos si la hipótesis es verdadera. Suponiendo que esta causa común aleatoria no guarda relación con la relación causal, tras la refutación de este método, los valores p obtenidos deberían ser relativamente altos, lo que respalda la hipótesis causal original.

Sin embargo, es importante señalar que un valor p grande sólo indica que la causa común aleatoria agregada tiene poca influencia en la estimación causal, pero no puede probar directamente que la hipótesis original sea correcta, ya que puede haber otros factores influyentes no considerados.

Después de utilizar el método de causa común aleatoria, el nuevo valor del efecto se muestra en la figura 2, y La relación entre el efecto original y el nuevo efecto se muestra en la figura 3.

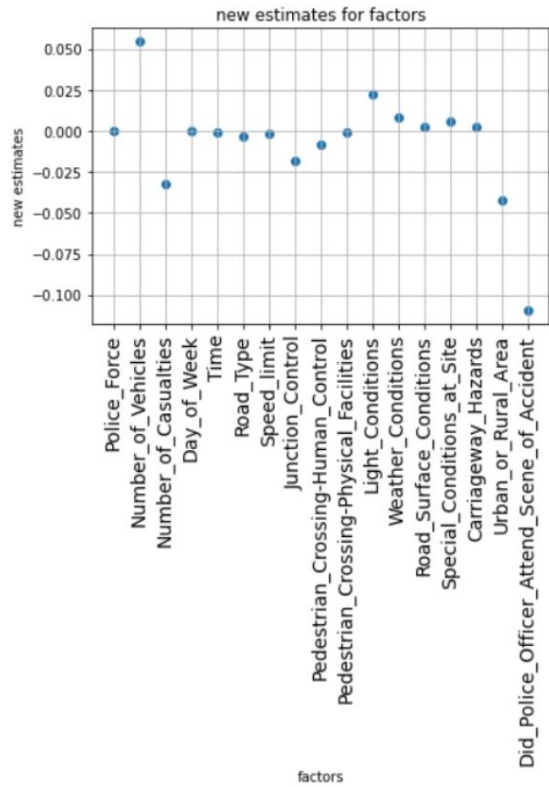


Figura 2. Valores reestimados.

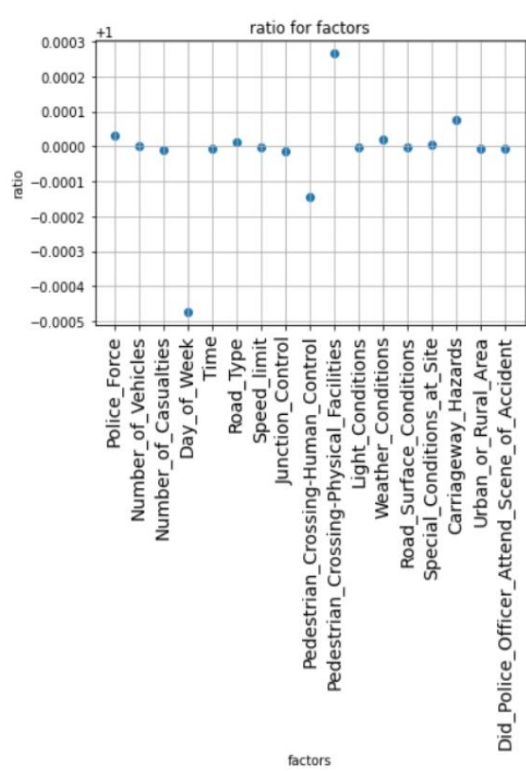


Figura 3. Valores de proporción.

La razón entre el efecto original y el nuevo efecto suele ser cercana a 1, lo que indica que la estimación causal no cambia mucho. Esto demuestra que la hipótesis original es correcta.

El valor P se muestra en la figura 4.

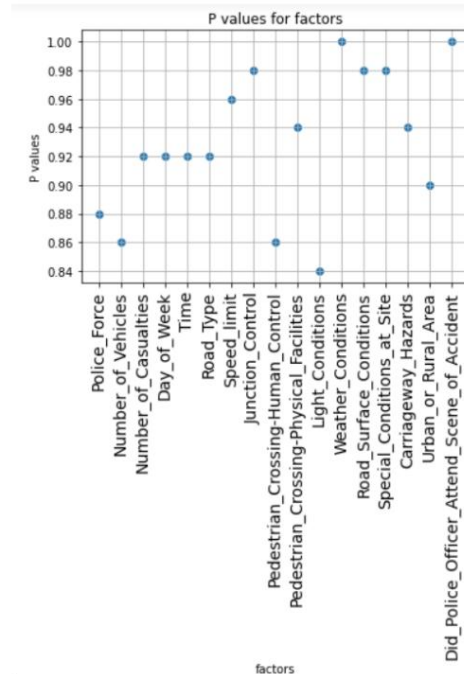


Figura 4. Valores p.

Como se muestra en los datos de la figura, los valores p son generalmente altos, y los valores p de las variables "¿Asistió el agente de policía al lugar del accidente?" y "Condiciones meteorológicas" alcanzan 1. Esto sugiere que la adición de causas comunes aleatorias prácticamente no afecta las relaciones causa-efecto. Por lo tanto, las relaciones causales en el gráfico causal obtenido en el segundo paso apenas se ven afectadas por las covariables. Esto refleja que la credibilidad de la relación causal entre estos diecisiete factores y la gravedad del accidente es muy alta.

### 3. Conclusión

Este estudio identifica las complejas relaciones causales entre todos los factores, entre los cuales las condiciones físicas del tránsito peatonal y las condiciones climáticas son las causas fundamentales que conducen a los accidentes de tránsito y al despliegue de las fuerzas policiales. En la estimación, este trabajo concluye que la intervención en el volumen de tráfico y la presencia policial tiene un gran impacto en la gravedad de los accidentes de tránsito. Además, este trabajo realiza una prueba de refutación utilizando el método de causa común aleatoria y métodos estadísticos, y los resultados muestran que la relación causal mencionada es altamente creíble. Por lo tanto, los resultados de la minería de relaciones causales y el experimento de intervención son de valor de referencia para la situación de tráfico. Controlar el flujo de tráfico y aumentar las patrullas policiales pueden ayudar eficazmente a reducir la ocurrencia de accidentes de tráfico.

Este artículo introduce por primera vez la inferencia causal en el campo del análisis de accidentes de tráfico. No solo realiza un estudio de la relación causal para revelar las razones que conducen a la gravedad del accidente, sino que también, desde la perspectiva de la mejora, realiza investigación. Esto puede proporcionar a futuros investigadores un nuevo método para estudiar los accidentes de tráfico. Al analizar la mejora de los accidentes de tráfico desde la perspectiva de la intervención, este estudio ayuda a los departamentos de gestión del tráfico a identificar prioridades de mejora y reducir la incidencia de accidentes de tráfico.

Este estudio se basa en la inferencia causal y solo identifica la relación causal general de todos los factores. En el futuro, también se podrán estudiar las relaciones causales específicas entre cada factor, a fin de mostrar la



relaciones causales de manera más profunda y específica, y utilizar esto para determinar la causa raíz del efecto de la intervención.

#### Referencias

- [1] XG Guo (2020). Investigación sobre la predicción de la gravedad de los accidentes de tráfico basada en la plataforma Spark. Universidad de Yunnan.
- [2] YL Luo (2020). Predicción del riesgo de tráfico vial basada en big data. Universidad de Changsha. Ciencia y tecnología.
- [3] L. Zheng, P. Gu y J. Lu (2021). Análisis de causas de accidentes de tráfico extremadamente graves basado en el árbol de fallos difusos ts y la red bayesiana. *Revista de Información y Seguridad del Transporte*, 39(4): 43-51+59.
- [4] TY Han, S. Tian, KG Lyu, X. Li, JT Zhang y L. Wei (2021). Análisis de redes sobre las causas de accidentes de tráfico graves basado en minería de texto. *China Safety Science Journal*, 31(9): 150-6.
- [5] E. Alomari, R. Mehmood e I. Katib (2019). Detección de eventos de tráfico vial mediante datos de Twitter, aprendizaje automático y Apache Spark, *IEEE SmartWorld 2019, Inteligencia y Computación Ubicuas, Computación Avanzada y Confiable, Computación y Comunicaciones Escalables, Computación en la Nube y Big Data, Internet de las Personas e Innovación en Ciudades Inteligentes (SmartWorld/SCALCOM/UIC/ATC/CBDCom/IOP/SCI)*, Leicester, Reino Unido, págs. 1888-95.
- [6] A. Saraswathi, A. Mummoorthy, A. Raman GR y KP Porkodi (2019). Sistema de monitoreo de tráfico en tiempo real usando Spark, *Conferencia Internacional sobre Tendencias Emergentes en Ciencia e Ingeniería (ICESE) de 2019*, Hyderabad, India, págs. 1-6.
- [7] YD Guo, XY Zhu, W. Guo y B. She (2020). Algoritmo paralelo para la estimación de la densidad del núcleo de la red vial basado en el marco de computación Spark. *Geomática y Ciencias de la Información de la Universidad de Wuhan*, 45(2): 289-95.
- [8] ZG Ma, XH Xu y XE Liu (2022). Tres marcos analíticos de inferencia causal y sus aplicaciones. *Revista China de Ingeniería*, 44(7): 1231-43.
- [9] Y. Zhang (2013). Modelo de inferencia causal en la evaluación cuantitativa de políticas educativas y las implicaciones de los métodos mixtos. *Investigación Educativa de la Universidad de Tsinghua*, 2013(3):29-40.
- [10] SY Li y AY Liu (2022). "Niños melancólicos": ¿La participación en escuelas intensivas genera emociones negativas? Inferencia causal basada en datos de la Encuesta de Panel Educativo Chino (CEPS). *Revista China de Sociología*, 42(2): 60-93.
- [11] XH Liu, HK Li, LJ Wang, AL Liu, Y. Qi, SS Sun, LF Zhang, HJ Ji, GY Liu, H. Zhao, YN Jiang, JY Li, CC Song, X. Yu, L. Yang, JC Yu, H. Feng, FJ Yang y FZ Xue (2022). Metodología de inferencia causal para el cribado de indicadores de salud. *REVISTA CHINA DE CONTROL Y PREVENCIÓN DE ENFERMEDADES*, 26(10): 1180-6.