






FraudAuditor: un enfoque de análisis visual para Fraude colusorio en seguros de salud

Jiehui Zhou , Xumeng Wang , Jie Wang, Hui Ye, Huanliang Wang, Zihan Zhou, Dongming Han , Haochao Ying , Jian Wu y Wei Chen 

Resumen—El fraude colusorio, en el que múltiples estafadores se confabulan para defraudar los fondos del seguro médico, amenaza el funcionamiento del sistema de salud. Sin embargo, los métodos estadísticos y basados en el aprendizaje automático existentes tienen una capacidad limitada para detectar el fraude en el escenario del seguro médico debido a la gran similitud de los comportamientos fraudulentos con las visitas médicas normales y la falta de datos etiquetados.

Para garantizar la precisión de los resultados de la detección, es necesario integrar conocimientos expertos en el proceso de detección de fraude. Al trabajar en estrecha colaboración con expertos en auditoría de seguros médicos, proponemos FraudAuditor, un enfoque de análisis visual de tres etapas para la detección de fraude colusivo en los seguros médicos. Específicamente, primero permitimos a los usuarios construir interactivamente una red de visitas conjuntas para modelar de manera integral las relaciones de visitas de diferentes pacientes. En segundo lugar, se ha diseñado un algoritmo de detección de comunidades mejorado que considera la intensidad de la probabilidad de fraude para detectar grupos fraudulentos sospechosos. Finalmente, a través de nuestra interfaz visual, los usuarios pueden comparar, investigar y verificar el comportamiento sospechoso de los pacientes con visualizaciones personalizadas que admiten diferentes escalas de tiempo. Realizamos estudios de casos en un escenario de atención médica del mundo real, es decir, para ayudar a localizar el grupo de fraude real y excluir el grupo de falsos positivos. Los resultados y los comentarios de los expertos demostraron la eficacia y usabilidad del enfoque.

Términos del índice : análisis visual, fraude colusorio, detección de fraude, seguro médico.

I. INTRODUCCIÓN

Un sistema de seguro médico EFICAZ desempeña un papel importante UN papel en la gestión de los recursos sanitarios, mejorando la vida calidad para las personas y mantener la estabilidad social. Más de 1.300 millones de personas se han inscrito en el Seguro Médico Básico Nacional en China¹. Sin embargo, el aumento del fraude en los seguros médicos

Manuscrito recibido el 28 de octubre de 2022; revisado el 13 de enero de 2023; aceptado el 14 de febrero de 2023. Fecha de publicación 27 de marzo de 2023; fecha de la versión actual 8 de mayo de 2023. Este trabajo fue apoyado por NSFC bajo las subvenciones 62132017 y 62202244. Recomendado para su aceptación por J. Choo, T. Ropinski e Y. Hu. (Autores correspondientes: Wei Chen; Haochao Ying).

Jiehui Zhou, Huanliang Wang, Zihan Zhou, Dongming Han y Wei Chen trabajan en el State Key Lab de CAD&CG, Universidad de Zhejiang, Hangzhou, Zhejiang 310027, China (correo electrónico: zhoujiehui@zju.edu.cn; wanghuanliang@zju.edu.cn; zhouzihan@zju.edu.cn; dongminghan@zju.edu.cn; chenvis@zju.edu.cn).

Xumeng Wang trabaja en TMCC, CS, Universidad de Nankai, Tianjin 300071, China (correo electrónico: wangxumeng@nankai.edu.cn).

Jie Wang trabaja en el Grupo Alibaba, Hangzhou, provincia de Zhejiang 311121, China (correo electrónico: siwei.wj@alibabainc.com).

Hui Ye trabaja en Tencent, Shenzhen, Guangdong 518054, China (correo electrónico: hazelye@tencent.com).

Haochao Ying trabaja en la Escuela de Salud Pública, Laboratorio Clave de Medicina Preventiva Inteligente de la Provincia de Zhejiang, Universidad de Zhejiang, Hangzhou, Zhejiang 310027, China (correo electrónico: haochaoying@zju.edu.cn).

Jian Wu trabaja en la Facultad de Medicina del Segundo Hospital Afiliado, Facultad de Salud Pública, Instituto de Wenzhou, Universidad de Zhejiang, Hangzhou, Zhejiang 310027, China (correo electrónico: wujian2000@zju.edu.cn).

Identificador de objetos digitales 10.1109/TVCG.2023.3261910

1http://en.nhc.gov.cn/2020-06/28/c_80923.htm

Los acontecimientos se han convertido en un grave problema social. Según la inspección realizada por la Administración Nacional de Seguridad Sanitaria y el Ministerio de Seguridad Pública de China, casi la mitad de 815.000 institutos de salud tuvieron costos de fondos inadecuados o incluso ilegales en 2020, lo que provocó una pérdida económica de más de 22.300 millones de yuanes (3.400 millones de dólares)². El fraude colusorio emergente es el más grave y urgente de estos acontecimientos [1]. Los estafadores se confabulan para comprar medicamentos con el reembolso del seguro y retirarlos en efectivo. La enorme cantidad de fraude trae graves consecuencias. Existe una necesidad urgente de métodos de detección eficientes y eficaces para identificar rápidamente el fraude colusorio y evitar mayores pérdidas.

La detección de fraude colusorio en los seguros médicos enfrenta dos desafíos. En primer lugar, es difícil distinguir el comportamiento de los estafadores en las visitas médicas del de los pacientes normales. Por lo general, los estafadores compran con frecuencia grandes cantidades de medicamentos fácilmente comercializables. Sin embargo, debido a la necesidad de mantener la medicación a largo plazo, los pacientes con enfermedades crónicas y aquellos que requieren tratamiento con medicina tradicional china (MTC) tienen comportamientos de compra similares a los de los estafadores. En segundo lugar, la auditoría manual es necesaria pero laboriosa. La identificación errónea es inaceptable para la detección de fraude porque un paciente tiene que asumir la responsabilidad legal después de ser reconocido como defraudador. Verificar el fraude requiere que los auditores sinteticen una gran cantidad de información contextual, como el monto del reembolso, el grado en que la enfermedad del paciente y los medicamentos coinciden y el tiempo de las visitas.

Los métodos existentes de detección de fraudes colusorios difícilmente pueden hacer frente a estos desafíos. Los métodos existentes se centran en modelar las relaciones entre defraudadores mediante gráficos y detectar grupos fraudulentos mediante métodos estadísticos o de aprendizaje automático (ML).

Los enfoques estadísticos utilizan características estructurales y de atributos [2], [3],[4] o análisis espectral[5] para detectar subestructuras anómalas (es decir, grupos/eventos de fraude). Sin embargo, los expertos en auditoría nos dijeron que estos métodos son propensos a dar falsos positivos, debido a la naturaleza ambigua del fraude colusorio en los seguros de salud. Excluir falsos positivos requiere mucho tiempo para los auditores y puede reducir significativamente la eficiencia de la detección. Los métodos de ML utilizan principalmente modelos de redes neuronales gráficas (GNN) para detectar fraudes colusorios [6], [7], [8]. Los estafadores y sus asociaciones se construyen como gráficos homogéneos o heterogéneos. Los GNN entrenados con datos etiquetados pueden generar la representación de estafadores y aplicarse aún más para juzgar a personas no etiquetadas. Desafortunadamente, grandes cantidades de datos etiquetados son indispensables para las GNN de alto rendimiento.

2http://en.ce.cn/main/latest/202102/22/t20210222_36327961.shtml

Sin suficiente fraude etiquetado, los modelos GNN no son aplicables en nuestro escenario.

Para abordar estos desafíos, proponemos un novedoso enfoque de análisis visual para ayudar a los expertos en auditoría de seguros médicos a identificar grupos sospechosos, investigar el comportamiento de visitas de pacientes sospechosos y validar los resultados del fraude colusorio. Proponemos una red de covisitas para representar la relación entre pacientes.

Los pesos de los bordes se calculan extrayendo las características de los estafadores colusorios, como el intervalo de tiempo y el número de visitas. Los grupos sospechosos con múltiples visitas simultáneas al mismo lugar pueden identificarse mediante un algoritmo de detección de comunidad ponderada. El algoritmo está integrado en un sistema prototipo, FraudAuditor, que ayuda a los expertos a explorar interactivamente y mejorar los resultados de detección de modelos.

FraudAuditor puede ayudar a los expertos a localizar y examinar rápidamente el fraude observando enlaces de visitas conjuntas en visualizaciones del comportamiento médico del paciente. En combinación con información contextual, como información sobre enfermedades, medicamentos y tarifas, se pueden verificar y excluir grupos de falsos positivos. Proporcionamos estudios de casos y entrevistas a expertos en escenarios reales de seguros de salud para validar la efectividad del enfoque propuesto.

Las contribuciones en este trabajo incluyen:

Una caracterización del problema que resume los requisitos de detección de fraude colusivo en el escenario de la salud seguro.

Un novedoso enfoque de análisis visual de tres etapas para detectar fraude colusorio en seguros de salud que considera el patrón de visitas de los grupos de fraude y el conocimiento de los expertos.

Un prototipo de sistema interactivo , FraudAuditor, para facilitar la identificación, examen y validación de grupos de fraude colusorios sospechosos.

II. TRABAJO RELACIONADO

A. Modelos de detección de fraude colusivo

Los modelos de detección de anomalías se aplican ampliamente para detectar fraude colusorio a partir de datos gráficos que registran eventos interpersonales mediante la identificación de grupos con comportamientos inesperados [3], [9], [10], [11], [12]. Los métodos relacionados se pueden dividir en modelos basados en estadísticas y métodos de ML.

Los modelos basados en estadísticas identifican anomalías a través de la información estadística de nodos, aristas o subgráficos. Por ejemplo, Akoglu et al. [2] extrajeron características estructurales, como el grado de nodo o la centralidad, del gráfico para encontrar egonets. SpamCom [3] identificó comunidades de spammers en Twitter mediante el uso de características de estructura y atributos como la similitud del contenido de Twitter, la topología del usuario y el perfil del usuario. En escenarios de atención sanitaria, Chen et al. [5] aplicaron un método de detección comunitaria basado en análisis de espectro para detectar casos de fraude en derivaciones de pacientes a partir de un gráfico bipartito de médicos y especialistas. Zhao y cols. [13] generó una red de información dinámica y heterogénea que contiene pacientes, hospitales y enfermedades. Luego, identificaron anomalías que se ajustan a patrones de fraude predefinidos (por ejemplo, el tratamiento único de alto costo) durante períodos fijos o variables. Los métodos basados en estadísticas pueden producir candidatos iniciales para el fraude, pero pueden tener resultados erróneos, lo que requiere una mayor validación por parte de expertos.

Los métodos de ML suelen utilizar GNN para detectar fraudes, ya que es potente para aprender una representación profunda de los nodos. Estudios anteriores se han realizado sobre gráficos homogéneos [6], [14] o heterogéneos [7], [8], [15]. Wang y cols. [6] construyeron una red de revisores en tiendas de aplicaciones en línea, donde los nodos (es decir, revisores) están conectados si han revisado la misma aplicación. Las revisiones y las características de comportamiento de los revisores se extraen de los registros de revisiones. Luego, se entrena y utiliza un modelo de red convolucional gráfico para detectar más estafadores en función del fraude identificado. Para detectar la colusión en préstamos de consumo fraudulentos por parte de personas con diversos roles (por ejemplo, vendedores e intermediarios), Xu et al. [7] proponen GRC, un modelo GNN novedoso, que aprende representaciones de diferentes tipos de individuos y detecta fraudes en préstamos mediante el uso de mecanismos de atención e imponiendo campos aleatorios condicionales.

Sin embargo, estos métodos de lavado de dinero están supervisados o semisupervisados y, por lo tanto, requieren datos etiquetados como fraude, algo que falta en nuestro escenario de seguro médico.

Dado que el límite entre el comportamiento fraudulento y el comportamiento normal en los seguros de salud puede no estar claro, los modelos automatizados difícilmente pueden aprender a juzgar correctamente y lograr una precisión satisfactoria. Por lo tanto, nuestro enfoque integra un modelo de detección basado en gráficos con una interfaz visual, que admite la exploración interactiva de datos, la optimización del modelo y la validación de resultados.

B. Enfoques de análisis visual para la detección de fraude

Para la detección de fraude con intervención humana, los estudios existentes emplean análisis visuales para ayudar a los usuarios a comprender e implementar tareas de detección desde la perspectiva de retratos individuales, cambios dramáticos y eventos interpersonales.

Los retratos individuales incluyen registros de alta dimensión, que pueden describirse y compararse mediante representaciones de glifos [16], [17], [18]. Los tres glifos circulares de TargetVue [17] representan las actividades de comunicación, características e interacciones sociales de los usuarios de Twitter. Los glifos yuxtapuestos permiten a los usuarios comparar los comportamientos de diferentes individuos y descubrir posibles fraudes, como los robots sociales. Para analizar y detectar patrones de fraude en transacciones bancarias, Macas et al. [18] ofrecieron diferentes glifos para caracterizar a los clientes del banco. Dependiendo del monto de la transacción, los beneficiarios y el tiempo de la transacción, el glifo tiene una forma circular o rectangular complementada con una serie de símbolos, que mejoran la comprensión del analista de los perfiles de transacciones típicos/atípicos.

Los cambios dramáticos también son un punto importante de las conductas fraudulentas. Estudios previos han diseñado múltiples técnicas de representación para visualizar información temporal, como visualización de secuencias [19], [20], diseños radiales [21], [22] y calendario [23], etc. FluxFlow [20] demuestra el Impacto de información anómala (p. ej., rumores) que se difunde a través de círculos de colores agrupados en una línea de tiempo. Bertini et al. [21] propusieron SpiralView, que utiliza gráficos de radar con ejes de tiempo en espiral para mostrar cómo las alertas cambian con el tiempo para detectar patrones periódicos sospechosos. TaxThemis [23] utiliza mapas de calor de calendario para mostrar evidencia de transferencia de ingresos a través de contribuyentes relacionados.

Los eventos interpersonales se pueden resumir mediante visualización gráfica. Por ejemplo, fraude financiero entre compradores y vendedores.

puede reflejarse en patrones estructurales anómalos compuestos de nodos y bordes [24]. Niu et al. [4] utilizó un diagrama de enlace de nodo para demostrar la red de garantía de préstamos, donde cada nodo pertenece a una comunidad definida por un algoritmo de caminata aleatoria y está codificado con el color correspondiente. Para identificar anomalías colectivas, Tao et al. [25] propusieron un gráfico de correlación de alto orden para respaldar los procesos de análisis que comienzan con un nodo anormal. Los nodos correspondientes que contribuyen a la anomalía se pueden identificar fácilmente mediante el gráfico de correlación de alto orden.

Nuestro sistema incorpora visualización de gráficos y secuencias. Para centrarse en el fraude colusorio en escenarios de seguros médicos, nuestro sistema proporciona información contextual más rica, como enfermedades, medicamentos y frecuencia de visitas.

III. CARACTERIZACIÓN DEL DOMINIO

A través de una colaboración intensiva con expertos en seguros médicos, obtenemos acceso a datos de seguros médicos del mundo real, aprendemos sobre los patrones de fraude colusorio y resumimos un conjunto de requisitos de diseño.

A. Descripción de datos

Los datos utilizados en este documento provienen de la Administración de Seguridad Sanitaria local con la que colaboramos. Bajo la guía de expertos, excluimos campos irrelevantes e información de identidad anónima. Dos tablas contienen los datos procesados: Tabla de visitas de pacientes: las filas representan las visitas de pacientes. Las columnas incluyen hora, identificación del paciente, institución médica, enfermedad diagnosticada y tarifa total. Por ejemplo, a las 16:23 del 12 de agosto de 2021, el paciente P1 acudió a un hospital. A P1 se le diagnosticó hipertensión y se le reembolsaron 36,35 yuanes por medicamentos a través del seguro médico.

Tabla de medicamentos: los médicos recetan medicamentos para cada visita del paciente. El nombre y la dosis de cada medicamento se registran en esta tabla. Tenga en cuenta que una receta puede incluir varios medicamentos. Por ejemplo, el médico le recetó fármacos antihipertensivos P1 , que consisten en perindopril y enalapril.

B. Especificación del problema

Durante el año pasado, trabajamos en estrecha colaboración con dos expertos en auditoría de seguros médicos con cuatro años de experiencia laboral. A través de múltiples entrevistas con ellos, aprendimos sobre el sistema de seguro médico, examinamos casos de fraude existentes y resumimos patrones de conductas de fraude colusorio.

En nuestro escenario, una parte del gasto médico puede reembolsarse cuando los pacientes asegurados pagan sus medicamentos. El fraude se produce cuando los pacientes asegurados venden los medicamentos en lugar de tomarlos después del reembolso. Para poder retirar dinero rápidamente, los estafadores necesitan recolectar una gran cantidad de medicamentos. Por lo tanto, los estafadores siempre se confabulan entre sí para comprar suficientes medicamentos en un corto período de tiempo. Para evitar el escrutinio, prefieren visitar clínicas o farmacias en zonas rurales o comunitarias mal reguladas. Sin embargo, el comportamiento de los grupos fraudulentos puede confundirse con el comportamiento normal de visita de los grupos de pacientes con enfermedades crónicas. Por ejemplo, si los médicos que tratan la enfermedad crónica atienden a los pacientes sólo en momentos específicos, existe una alta probabilidad de conductas de visita y compra de medicamentos

ciertos pacientes irán al mismo lugar a una hora similar. Por lo tanto, los pacientes normales también pueden tener características de conexiones espacio-temporales y acciones grupales similares al grupo de fraude colusorio, es decir, los pacientes pueden visitar ciertos institutos médicos juntos en ciclos similares. Para evitar errores de juicio, es indispensable que los expertos revisen y validen los grupos sospechosos haciendo referencia a la información contextual del comportamiento de las visitas de los pacientes.

Para comprender mejor las necesidades de los expertos en auditoría a la hora de detectar y analizar el fraude colusorio, los entrevistamos y resumimos el proceso de auditoría actual en tres pasos.

1. Los expertos utilizan el sistema de auditoría (una interfaz gráfica de la base de datos) para conocer las características generales de los datos del seguro médico, como el costo de las reclamaciones. Luego, reducen el alcance de la investigación filtrando a los pacientes de interés, como aquellos cuyos costos superan los 10.000 yuanes. Dado que los estafadores pueden tener múltiples trucos, los expertos requieren intentos repetidos para evitar omisiones. Luego, los expertos aprovechan las reglas manuales para filtrar a los pacientes e identificar mejor los grupos sospechosos. Por ejemplo, los expertos pueden identificar pacientes con un número inusualmente alto de compras de medicamentos durante un período de tiempo o separar secuencialmente grupos con una gran superposición de lugares de visita y tiempos de visita similares (por ejemplo, un grupo de cinco pacientes visita con frecuencia una farmacia específica dentro de 1 h para compra de medicamentos).
2. Los expertos exploran la lista de grupos sospechosos para comenzar con los grupos con peligros potenciales graves y aprovechar la oportunidad para detener las pérdidas a tiempo. El peligro potencial de un grupo se puede estimar según el tamaño del grupo y el gasto total de la reclamación. Además, los expertos también pueden fusionar grupos con características similares durante la navegación para mejorar la eficiencia del análisis.
3. Los expertos juzgan si un grupo es un fraude colusorio o un falso positivo examinando los detalles de los comportamientos de visita de los pacientes (enfermedades, medicamentos, hospitales, etc.). Se investigarán más a fondo los grupos altamente sospechosos (p. ej., inspecciones por vídeo de vigilancia y entrevistas telefónicas o in situ con los pacientes).

C. Análisis de requisitos

Como se mencionó anteriormente, el proceso de auditoría requiere una gran cantidad de inspecciones manuales, que requieren mucho tiempo y trabajo. Basándonos en conversaciones con expertos, resumimos los requisitos en tres niveles.

- Los usuarios necesitan conocer las estadísticas de los pacientes y sus conexiones de comportamiento desde el nivel de visión general .
- R1 Mostrar distribución de atributos de registros médicos: una descripción general permite a los usuarios comprender el conjunto de datos y encontrar un punto de partida para detectar fraude. Por ejemplo, los usuarios pueden conocer un rango de gastos razonable a partir de la distribución de los gastos de los pacientes. Luego, se deben revisar los pacientes cuyos gastos superen el umbral.
 - R2 Permitir un filtrado de datos flexible: nuestro conjunto de datos incluye registros de visitas de muchos pacientes. Algunos de ellos no necesitan ser auditados debido a pequeños gastos o un número limitado de visitas. Filtrar pacientes según condiciones apropiadas especificadas por el usuario puede mejorar la eficiencia del análisis.
 - R3 Identificar las conexiones conductuales entre pacientes: Conexiones de conductas de visita y compra de medicamentos

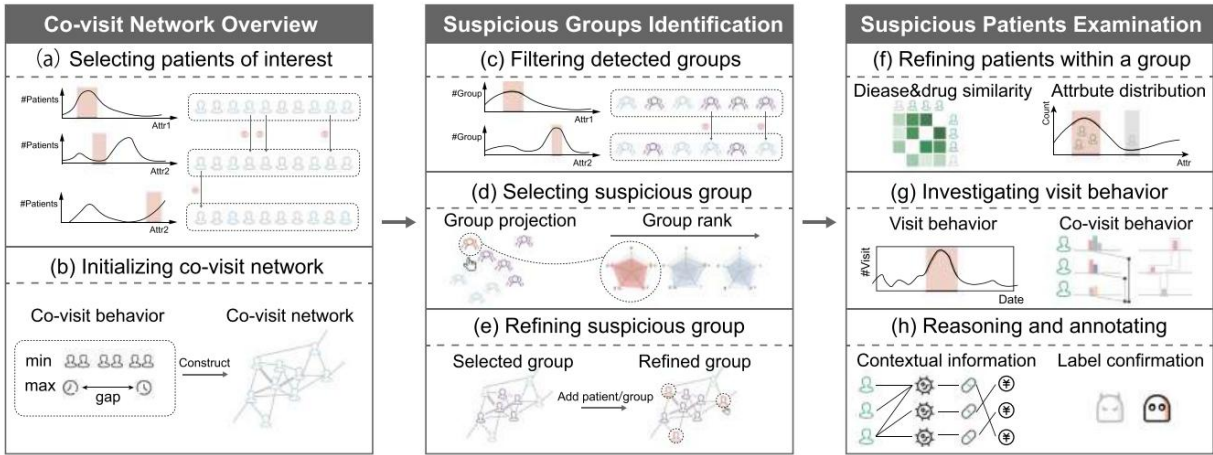


Fig. 1. El enfoque de tres etapas que puede ayudar a los usuarios a identificar, examinar y verificar grupos sospechosos de fraude colusorio en seguros de salud.

Los comportamientos son la base para detectar grupos de fraude. Por lo tanto, estas conexiones deben identificarse según el conocimiento experto, es decir, restricciones especificadas por el usuario. Por ejemplo, se considera que los pacientes están potencialmente asociados sólo si visitan el mismo lugar en menos de 15 minutos más de cinco veces.

Se deberían implementar más procesos de detección de fraude a nivel de grupo .

R4 Detectar grupos de pacientes: los grupos de pacientes se pueden detectar en función de varias reglas especificadas por el usuario (por ejemplo, si existen conexiones de comportamiento específicas o si el gasto total excede un límite). Se puede aprovechar la automatización para garantizar la eficiencia de la detección de grupos.

R5 Admite la selección de grupos sospechosos: dada una gran cantidad de grupos candidatos detectados, los usuarios deben localizar rápidamente aquellos con mayor sospecha o peligro. Recomendar grupos utilizando enfoques de clasificación eficaces puede acelerar la selección de grupos.

Finalmente, los usuarios deben verificar los detalles del grupo desde el nivel del paciente y encontrar evidencia de fraude.

R6 Apoyar la verificación de pacientes sospechosos: comprender las similitudes intragrupo de los comportamientos de los pacientes, como enfermedades prescritas, compras de medicamentos y selección de instituciones médicas, puede ayudar a los usuarios a excluir pacientes irrelevantes y examinar a los pacientes sospechosos.

R7 Visualizar los registros de visitas de un paciente individual: como se mencionó anteriormente, la detección automática difícilmente puede diferenciar los grupos fraudulentos de los pacientes con necesidades de visitas específicas, lo que genera falsos positivos. Los usuarios deben examinar a los estafadores sospechosos identificados para probar o refutar sus sospechas. Visualizar los historiales de visitas médicas de los pacientes podría ayudar a los usuarios a recopilar evidencia sobre la continuidad y racionalidad de las visitas. De esta forma se pueden diferenciar grupos fraudulentos y grupos de falsos positivos basándose en el conocimiento experto.

IV. NUESTRO ENFOQUE

Esta sección proporciona una descripción general del análisis visual. enfoque y presenta los dos modelos empleados.

A. Descripción general del enfoque

Con base en los requisitos de diseño elaborados en la Sección III, proponemos un enfoque de análisis visual (ver Fig. 1), que permite a los usuarios analizar registros de seguros médicos en múltiples niveles para profundizar en los datos de interés, localizar grupos fraudulentos sospechosos y encontrar evidencia a nivel de paciente para verificar el fraude colusorio.

Nuestro enfoque consta de tres etapas: (1) descripción general de la red de visitas conjuntas, (2) identificación de grupos sospechosos y (3) examen de pacientes sospechosos.

Descripción general de la red de visitas conjuntas: en la primera etapa, los usuarios buscan una comprensión general de los datos verificando las distribuciones de atributos (R1, Fig. 1-(a)). Según la distribución de los datos y el conocimiento del dominio, los usuarios filtran a los pacientes para su análisis (R2, Fig. 1-(a)). A continuación, los usuarios verifican las conexiones entre los pacientes filtrados de forma interactiva (R3, Fig. 1-(b)). Los comportamientos colusorios pueden revelarse por los intervalos de tiempo de las visitas o el número de covisitas. Se ayuda a los usuarios a especificar la definición de comportamientos cómplices estableciendo umbrales para los intervalos de tiempo y el número de covisitas.

Identificación de grupos sospechosos: en la segunda etapa, nuestro sistema emplea un método de minería de grupos (consulte la Sección IV-B) para detectar grupos cómplices de acuerdo con la definición especificada por el usuario (R4). Luego, nuestro sistema proporciona múltiples estrategias de selección para ayudar a los usuarios a localizar grupos objetivo de la lista de grupos detectados (R5). Las estrategias viables son el filtrado de atributos múltiples (Fig. 1-(c)), la comparación de grupos y la clasificación de grupos (Fig. 1-(d)). Los usuarios también pueden agregar pacientes o grupos vecinos para optimizar los resultados de detección (Fig. 1-(e)).

Examen de pacientes sospechosos: en la tercera etapa, nuestro sistema calcula la similitud de las enfermedades y medicamentos prescritos entre cada par de pacientes de un grupo (ver Sección IV-C). Según la similitud, los usuarios pueden evaluar la probabilidad de fraude colusorio (R6, Fig. 1-(f)). Los pacientes con baja probabilidad se pueden excluir de forma interactiva. A continuación, los usuarios pueden investigar a los pacientes en reposo inspeccionando sus comportamientos de visita en diferentes granularidades de tiempo (R7, Fig. 1-(g)). Nuestro sistema permite a los usuarios comprender rápidamente los períodos de tiempo y la frecuencia de las covisitas entre ellos. También proporcionamos información contextual, incluida la enfermedad, el medicamento y la tarifa, para ayudar a los usuarios a razonar y anotar si el comportamiento sospechoso es un fraude colusorio (R7, Fig. 1-(h)).

TABLA I
DEFINICIONES DE NOTACIÓN

Notation	Description
P	The patients set
<i>m</i>	The number of patients
V	The visits set
<i>n</i>	The number of visits
<i>t_i</i>	The time of visit <i>v_i</i>
<i>θ₁</i>	The maximum time gap for a co-visit
<i>θ₂</i>	The minimum number of co-visits
CV (<i>p_i</i> , <i>p_j</i>)	The co-visit behaviors between patient <i>p_i</i> and <i>p_j</i>
<i>w</i> (<i>v_i</i> , <i>v_j</i>)	The weight of a co-visit about visit <i>v_i</i> and <i>v_j</i>
<i>w</i> (<i>p_i</i> , <i>p_j</i>)	The weight of co-visits between patient <i>p_i</i> and <i>p_j</i>
W	The weight between patients
D	The diseases set
C	The number of visits for each disease in D
<i>w</i> (<i>d_i</i>)	The contribution of disease <i>d_i</i> to the similarity of patients <i>p_i</i> and <i>p_j</i>
<i>c_i</i>	The number of visits of disease <i>d_i</i>
<i>sim</i> (<i>p_i</i> , <i>p_j</i>)	The similarity of patient <i>p_i</i> and <i>p_j</i>

B. Minería grupal sospechosa

Para detectar grupos sospechosos con conexiones espacio-temporales y características de acción grupal (ver Sección III-B), proponemos un método de minería grupal sospechoso para detectar fraude colusorio en seguro médico (R4). Nuestro método primero construye una red de visitas conjuntas. para representar la relación espacio-temporal entre los pacientes. Basado en la red de covisitas, el método utiliza una modularidad Algoritmo de detección de comunidades basado en optimización para extraer grupos sospechosos. Para mayor claridad en la descripción, hemos enumerado los notaciones en la Tabla I. Consulte el Algoritmo 1 para el pseudocódigo.

Construcción de red de covisitas: pacientes en un fraude colusorio El grupo visita con frecuencia la misma institución médica en períodos de tiempo relativamente cortos. Considerando tal característica, construir una red de covisitas G entre pacientes para resumir la Comportamientos de co-visita y detección de fraude colusorio. Un nodo en la red El trabajo representa a un paciente. Una ventaja entre los registros de dos pacientes los comportamientos de covisita entre los dos pacientes. si el medico instituciones de las dos visitas correspondientes de dos pacientes son las mismo, y el intervalo de tiempo es menor que un umbral θ_1 (el valor predeterminado es 1 h, que puede ajustarse a 6, 12 o 24 horas), se considera una co-visita. Para los pacientes p_i y p_j , sus comportamientos de covisita son representado como $CV(p_i, p_j) = \{(v_1, v_1), \dots, (v_s, v_s)\}$ y s es el número total de visitas que hicieron juntos.

Cálculo del peso del borde: El peso del borde indica el probabilidad de que los dos pacientes pertenezcan al mismo grupo. Nosotros calculó los pesos del borde $w(p_i, p_j)$ en función del número de co-visitas y el intervalo de tiempo de visita. Como se muestra en (1), el El peso de una covisita es inversamente proporcional al tiempo de la visita. brecha. Para evitar el impacto de visitas ocasionales con poco tiempo brecha en el peso, inspirado en la función de activación ReLU, nosotros establecer el tiempo límite en 10 minutos según la experiencia de los expertos, y los pesos inferiores a ese intervalo se consideran iguales.

$$w(v_i, v_j) = \frac{1}{\max(10 \text{ minutos}, |t_i - t_j|)}$$

$$0 \leq |t_i - t_j| \leq \theta_1$$
$$\text{de lo contrario}$$

(1)

Algoritmo 1: Minería grupal sospechosa.

Entrada: P: el conjunto del paciente; V: los registros de visitas; W: el peso entre pacientes; θ_1 : el intervalo de tiempo máximo; θ_2 : los tiempos mínimos de covisita.

Producto: G: la red de covisitas; SG: el sospechoso grupos.

1: $CV \leftarrow$ extraer comportamiento de covisita de V

2: $W \leftarrow 0$

3: para cada par de pacientes (p_i, p_j) haga

4: $w(p_i, p_j) \leftarrow 0$

5: si $|CV(p_i, p_j)| \geq \theta_2$ entonces

6: para cada visita conjunta (v_{ik}, v_{jk}) en $CV(p_i, p_j)$ haga

7: $w(v_{ik}, v_{jk}) = \frac{1(|t_i - t_j| \leq \theta_1) \max(10 \text{ minutos}, |t_i - t_j|)}{w(p_i, p_j) + w(v_{ik}, v_{jk})}$

8: 9: final para

10: terminar si

11: fin para

12: $SOL \leftarrow (P, CV, W)$

13: $SG \leftarrow \text{CDLIB.ALGORITMOS.LOUVAIN}(G, W)$

14: regresar G, SG

El peso del borde $w(p_i, p_j)$ entre dos pacientes es el total de sus ponderaciones de covisitas, definidas como (2). Un umbral ajustable Aquí se establece θ_2 (por defecto 4) para el número mínimo de covisitas. para evitar factores aleatorios. El peso de la covisita es menor que el umbral indica la baja probabilidad de que ambos pertenezcan al Mismo grupo.

$$w(p_i, p_j) = \frac{\sum_{z=1}^{|CV(p_i, p_j)|} w(v_{iz}, v_{jz})}{|CV(p_i, p_j)| \geq \theta_2 \text{ de lo contrario}}$$

(2)

Detección comunitaria: para minar grupos sospechosos de la red de co-visita utilizamos Lovaina [26], una comunidad Algoritmo de detección basado en optimización de modularidad. El algoritmo es aplicable a gráficos ponderados y admite la exclusión. de nodos no comunitarios, que pueden producir resultados de detección claros ya que la mayoría de los pacientes en el escenario sanitario son normales.

C. Similitud entre enfermedades y fármacos

Para verificar grupos de fraude colusorios, necesitamos calcular el similitud entre pacientes según sus enfermedades prescritas y medicamentos correspondientes (R6). Al principio intentamos calcular el similitud según los textos de cadena de enfermedades y drogas, pero los resultados no fueron satisfactorios. Por ejemplo, tal cálculo llevaría a que el dolor de cabeza fuera similar al dolor de estómago y no es similar al derrame cerebral, cuando en realidad tanto los dolores de cabeza como los derrames cerebrales Son trastornos cerebrales con una relación más estrecha. Más tarde, encontramos que las enfermedades o los medicamentos tienen una codificación jerárquica, lo que puede reflejar la información de similitud. La codificación ICD10 de ³ enfermedades y la codificación estándar de medicamentos codifica enfermedades y drogas jerárquicamente por clases grandes, medianas y pequeñas.

3<https://en.wikipedia.org/wiki/ICD-10>

4<https://code.nhsa.gov.cn/toDetail.html?infold=5546&CatalogId=2>

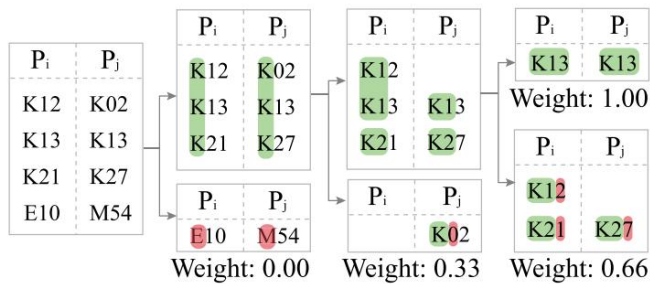


Fig. 2. Un ejemplo de cálculo del peso de la enfermedad.

Por ejemplo, las enfermedades J11 (influenza) y J18 (neumonía) son similares, pero son muy diferentes de M54 (dolor de espalda).

Por lo tanto, proponemos un método de cálculo de similitud basado en la coincidencia más cercana que considera los códigos de enfermedad/fármaco y el número de visitas para las enfermedades correspondientes. Para cada enfermedad/fármaco, necesitamos encontrar el más similar en el conjunto de enfermedades/fármacos de otro paciente, de modo que no se vea afectado por el orden específico de visitas. Supongamos que a un paciente le han prescrito varias enfermedades $D = \{d_1, d_2, \dots, d_l\}$. El número de visitas correspondiente a cada enfermedad es $C = \{c_1, c_2, \dots, c_l\}$.

Como se muestra en la Fig. 2, de la primera letra de las enfermedades de p_i y p_j , podemos ver que ambas han sido tratadas por enfermedades que comienzan con "K". Sin embargo, el E10 de p_i (es decir, diabetes tipo 1) y el M54 de p_j (es decir, dolor de espalda) no se comparten entre sí. Por lo tanto, E10 y M54 no aportan ninguna similitud de enfermedad.

A continuación, según el código de la segunda letra, el conjunto de enfermedades que comienza con K se puede dividir en tres conjuntos de clase media: K0, K1 y K2. La enfermedad K02 de p_j no corresponde a las enfermedades restantes de p_i , mientras que las enfermedades restantes se calculan según las letras de la tercera letra hasta que las enfermedades de los dos pacientes del conjunto sean exactamente iguales.

Cuanto más cercanas estén las dos enfermedades, mayor será su contribución al cálculo de la similitud. Por tanto, el peso de cada enfermedad viene determinado por el prefijo más largo del conjunto en el que permanece por última vez, teniendo

$$= \text{la} \frac{\text{la longitud de la letra del prefijo } w(d_i)}{\text{longitud total de la letra codificada}} \quad (3)$$

Por ejemplo, el peso de la enfermedad K02 es $1/3 \approx 0,33$, K12 es $2/3 \approx 0,66$ y K13 es 1,00. Lo mismo ocurre con las drogas. El número de visitas médicas por enfermedades/medicamentos también refleja su similitud. La similitud entre dos pacientes se calcula de la siguiente manera:

$$\text{sim}(p_i, p_j) = \frac{w(p_i) \cdot c_{p_i} + w(d_{p_j}) \cdot c_{d_{p_j}}}{w(p_i) \cdot c_{p_i} + c_{d_{p_j}}} \quad (4)$$

V. DISEÑO DEL SISTEMA

Para ayudar a los usuarios a implementar el enfoque mencionado en la Sección IV, desarrollamos un sistema de creación de prototipos interactivo, FraudAudiator. Esta sección presenta una descripción general del sistema e introduce los detalles del diseño visual y las interacciones.

A. Descripción general del sistema

El sistema contiene cuatro vistas, como se muestra en la Fig. 3: la vista de análisis de red, la vista de comparación de grupo, la vista de comparación de pacientes y la vista de comportamiento del paciente. Describimos un flujo de análisis para demostrar cómo estas cuatro vistas ayudan al usuario a descubrir, analizar y validar grupos sospechosos de fraude colusorio basados en datos de seguros médicos. El usuario primero aprende la distribución de datos del gráfico de barras de los atributos del paciente en la vista de análisis de red (R1), en función del cual puede filtrar interactivamente los datos de interés (R2). Luego, establece parámetros sobre el comportamiento de visitas conjuntas y genera la red de visitas conjuntas entre pacientes en la vista de red de visitas conjuntas de pacientes (R3). Los resultados del modelo de detección automática también se muestran en la red en tiempo real resaltando. La distribución de atributos, la similitud y la clasificación de los grupos detectados se pueden ver en la vista de comparación de grupos (R5). Hace clic en el grupo mejor clasificado y su posición en la red se resalta simultáneamente. En la vista de comparación de pacientes, compara diferentes pacientes del grupo utilizando la matriz de similitud de enfermedades y fármacos, el gráfico de barras apiladas y el gráfico de áreas (R6). A partir de ahí, selecciona varios pacientes sospechosos y pasa a la vista de comportamiento del paciente para realizar una investigación más exhaustiva.

Analiza visualmente el patrón de visitas y la distribución de visitas conjuntas de los pacientes a partir de la visualización de la secuencia de visitas. Combinado con la visualización de información contextual como enfermedades y medicamentos, infiere y etiqueta a estos pacientes como involucrados en fraude colusorio (R7).

B. Vista de análisis de red

La vista de análisis de red (Fig. 3(a)) tiene dos partes: (1) La vista de atributos del paciente brinda una descripción general de los pacientes al mostrar distribuciones de atributos y admite el filtrado interactivo de datos de interés (R1, R2). (2) La red de visitas conjuntas de pacientes permite a los usuarios definir de forma interactiva comportamientos de visitas conjuntas, explorar la red de visitas conjuntas resultante e inspeccionar visualmente los grupos sospechosos detectados por algoritmos automatizados (R3).

En la vista de atributos del paciente (Fig. 3(a1)), las barras indican la distribución de los atributos del paciente, incluida la distribución de los pacientes en términos de número de visitas, edad y tarifa total, así como el número de visitas a diferentes instituciones médicas.

Al principio, se seleccionan todos los pacientes y los usuarios pueden hacer clic en una barra para deseleccionar o volver a seleccionar los pacientes correspondientes. Los pacientes que no están seleccionados están representados por un fondo translúcido, y al pasar el mouse sobre la barra correspondiente aparecerá una información sobre herramientas que muestra el número total de pacientes que pertenecen a los pacientes originales y actuales en el intervalo, lo que facilita a los usuarios comparar la Distribución de pacientes bajo diferentes condiciones de filtrado.

En la red de visitas conjuntas de pacientes (Fig. 3(a2)), el panel de control anterior permite a los usuarios configurar de forma interactiva la definición del comportamiento de visitas conjuntas. Un control deslizante controla el número mínimo de visitas conjuntas y un menú desplegable establece el intervalo de tiempo máximo, como 1 h, 6 horas, etc. Los usuarios pueden explorar la red de visitas conjuntas haciendo clic en "Generar gráfico" y cambiar iterativamente la definición de covisita o las condiciones de filtrado en la vista de atributos del paciente si los resultados no son satisfactorios. Los grupos sospechosos detectados por el algoritmo automatizado se muestran en la red



Fig. 3. FraudAuditor facilita la identificación, examen y anotación de fraude colusorio en seguros de salud. (a) La vista de análisis de red admite el filtrado interactivo de los atributos del paciente y la construcción de una red de visitas conjuntas. (b) La vista de comparación de grupos proporciona filtrado interactivo, análisis de similitud y clasificación de grupos. (c) La vista de comparación de pacientes ayuda a los usuarios a analizar la similitud y distribución de enfermedades, medicamentos y otros atributos entre los pacientes dentro de un grupo y ayuda a seleccionar pacientes sospechosos para ser analizados. (d) La vista del comportamiento del paciente respalda la inspección y anotación de registros detallados de visitas de pacientes y el análisis de comportamientos de co-visita.

simultáneamente y marcados en violeta, donde se pueden filtrar grupos pequeños ajustando el control deslizante de tamaño mínimo de componente. Se aplicó el diagrama de enlace de nodo para visualizar el gráfico. Porque el diagrama de enlace de nodo funciona bien al mostrar relaciones directas e indirectas entre pacientes, lo que puede respaldar tareas de análisis de conexiones y exploración de topología. Cada nodo de la red representa un paciente y el borde entre dos nodos refleja su relación de covisita, cuyo ancho representa la fuerza de la relación de covisita (consulte la Sección IV- B).

La red admite zoom y desplazamiento para la navegación.

Los pacientes se resaltan y la identificación del paciente y la identificación del grupo (si corresponde) se muestran cuando el mouse se coloca sobre un nodo. Si el nodo pertenece a un grupo, se resaltan todos los nodos de ese grupo. Para ayudar a los usuarios a comprender la procedencia del análisis, los nodos seleccionados, sospechosos, normales y otros se asignan a diferentes codificaciones visuales.

C. Vista de comparación de grupos

La vista de comparación de grupos (Fig. 3-(b)) consta de tres partes: (1) La vista de atributos de grupo proporciona una descripción general de los atributos a nivel de grupo, así como capacidades de filtrado interactivo. (2) La proyección de grupo admite el análisis de similitud de grupos (3) La clasificación de grupo permite clasificar grupos en múltiples dimensiones. Mediante un filtrado inicial y una selección cuidadosa, los usuarios pueden identificar grupos sospechosos en los que deben centrarse para su análisis (R5).

La vista de atributos de grupo (Fig. 3-(b1)) utiliza gráficos de barras para mostrar la distribución de grupos en varias métricas, incluida la cantidad de pacientes (p), la tarifa total per cápita (f), la cantidad de visitas conjuntas. (c), el número promedio de días entre múltiples covisitas (d) y el intervalo de tiempo mínimo dentro de una covisita (g), que son métricas comunes utilizadas por los expertos en auditoría de seguros médicos para evaluar qué tan sospechoso es un grupo. La vista también admite el filtrado interactivo para ayudar a limitar los grupos que se analizarán.

La proyección de grupo (Fig. 3-(b2)) asigna grupos a un plano bidimensional, lo que ayuda a los usuarios a comparar similitudes y diferencias entre grupos y a detectar grupos o valores atípicos que merecen análisis. Utilizando las características a nivel de grupo mencionadas anteriormente, los grupos originales se representan como un conjunto de vectores de características. Para comparar la distribución de grupos sobre diferentes características, utilizamos kernel PCA [27] porque mantiene la covarianza de los datos y es capaz de manejar casos linealmente indistinguibles. En el resultado de la proyección, cada triángulo representa un grupo y la distancia entre ellos refleja su similitud hasta cierto punto. Dependiendo del estado del grupo, por ejemplo, filtrado o seleccionado por el usuario, se aplica una codificación visual diferente al grupo.

El grupo que necesita más investigación se puede seleccionar haciendo clic en él.

La vista de clasificación del grupo (Fig. 3-(b3)) proporciona un menú desplegable para seleccionar la palabra clave de clasificación, donde se pueden utilizar métricas de un solo grupo o puntuaciones generales. Creamos un gráfico de radar personalizado para comparar visualmente grupos según múltiples criterios. En

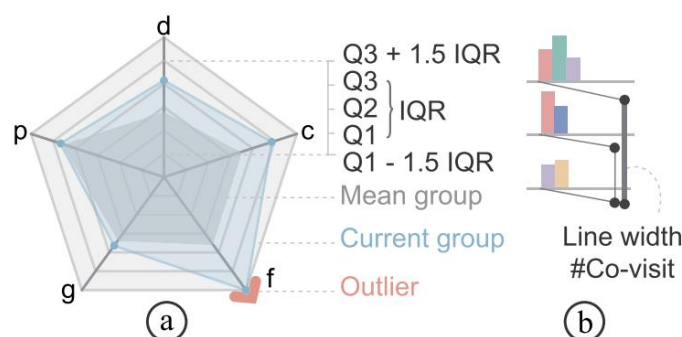


Fig. 4. Explicación del diseño visual. (a) Carta de radar personalizada. Cinco ejes representan diferentes métricas de grupo. El área gris más oscura representa el promedio de todos los grupos en estas métricas y el área azul representa el grupo actual. Si el valor en un eje excede $Q3 + 1.5 IQR$, se marca con una flecha como valor atípico. (b) El vínculo de co-visita de la metáfora de la nota, y el ancho representa el número de co-visitas.

Fig. 4-(a), cada eje representa la métrica mencionada anteriormente.

Cinco líneas grises atraviesan cada eje de adentro hacia afuera, representando estadísticamente el límite inferior, el primer, segundo, tercer cuartil y el límite superior de cada grupo en cada métrica. A modo de comparación, las transformaciones de datos (por ejemplo, agregar signos negativos) hacen que todas las métricas sean más anómalas a medida que crecen. Un área gris indica el valor promedio de cada métrica como referencia. El área azul muestra el grupo actual, cuyo tamaño refleja visualmente el grado de sospecha del grupo. Una flecha especial alerta a los usuarios sobre valores atípicos por encima de la valla superior. Al pasar el cursor sobre este glifo se muestran detalles del grupo y de cada métrica. Los usuarios pueden hacer clic para seleccionar el grupo correspondiente para su posterior análisis.

D. Vista de comparación de pacientes

La vista de comparación de pacientes (Fig. 3-(c)) tiene dos tipos de visualizaciones:

- (1) Una matriz de similitud de enfermedades y medicamentos que respalda las comparaciones entre pacientes y ayuda a los usuarios a determinar si las enfermedades y los medicamentos se corresponden entre sí.
 - (2) Gráficos de barras apiladas y gráficos de áreas del atributo del paciente para ayudar a analizar la contribución de diferentes pacientes al grupo.
- A través de similitudes y comparación de atributos, los usuarios pueden identificar pacientes sospechosos (R6).

Para seleccionar aún más a los pacientes sospechosos y excluir a los inocentes, los usuarios deben estudiar la similitud de las enfermedades y los medicamentos dentro de un grupo. Mostramos la similitud entre cada par de pacientes en una matriz (Fig. 3-(c1)). Las celdas se dividen en dos categorías: la esquina superior izquierda representa la similitud de medicamentos y la esquina inferior derecha representa la similitud de enfermedades. Tanto el eje horizontal como el vertical de la matriz son los pacientes dentro del grupo seleccionado y tienen el mismo orden. Cada celda está codificada por colores en verde con el grado de similitud entre los pacientes correspondientes; cuanto más oscuro, más parecido. Cuando el mouse pasa sobre una celda, se muestran los valores específicos de enfermedades y similitud de medicamentos y las identificaciones de pacientes correspondientes, y la celda diagonalmente simétrica también se resalta para comparar. Para descubrir patrones de agrupación de pacientes según enfermedades y medicamentos, el sistema admite el reordenamiento matricial, donde los niveles jerárquicos

Los métodos de agrupación se utilizan para determinar los pedidos de nuevos pacientes. El paciente puede ser seleccionado para un análisis más detallado haciendo clic en las etiquetas. Dado que normalmente hay como máximo una docena de estafadores en un grupo, la matriz es menos propensa al desorden visual.

Si se encuentran pacientes sospechosos (por ejemplo, con enfermedades similares pero grandes diferencias en los medicamentos), los usuarios necesitan información contextual para evaluar más a fondo la racionalidad del comportamiento de compra de medicamentos. En el gráfico de barras apiladas y el gráfico de áreas apiladas (Fig. 3-(c2)), los usuarios pueden estudiar información contextual sobre enfermedades, medicamentos, institutos médicos visitados, tarifas totales, etc., para verificar pacientes sospechosos. Los datos correspondientes a los pacientes seleccionados por los usuarios se resaltan, mientras que los datos de los pacientes no seleccionados se vuelven translúcidos para proporcionar una imagen clara de la proporción de estos pacientes en el grupo general en diferentes atributos. Cuando el mouse pasa sobre una barra o área, se muestra información sobre cada paciente seleccionado y todo el grupo en ese intervalo de atributos en una información sobre herramientas.

E. Vista del comportamiento del paciente

La vista del comportamiento del paciente (Fig. 3-(d)) contiene: (1) un gráfico de líneas del número de visitas a lo largo del tiempo para localizar anomalías y navegación; y (2) una visualización de la secuencia de visitas de los pacientes que muestra la evolución de las visitas de los pacientes y destaca las co-visitas. Los usuarios pueden analizar el comportamiento de las visitas en diferentes granularidades temporales y combinar información contextual rica y conocimiento del dominio para razonar y verificar si se trata de fraude colusorio (R7).

El gráfico de líneas (Fig. 3-(d1)) en la parte inferior refleja el cambio temporal en el número de visitas para ayudar a los usuarios a localizar períodos de tiempo con fluctuaciones significativas y detectar visitas periódicas, etc. Los cuadros grises representan períodos de tiempo seleccionados y permiten para deslizar y desplazar rangos.

En la visualización de la secuencia de visitas (Fig. 3-(d2)), cada línea de tiempo corresponde al historial de visitas de un paciente durante el período de tiempo seleccionado. Diferentes períodos de tiempo dan lugar a diferentes granularidades de visualización, como mes, semana, día, etc., lo que ayuda a mejorar la capacidad de lectura y reducir la carga cognitiva. Las barras en la línea de tiempo representan el comportamiento de visita del paciente, donde la posición codifica el tiempo de la visita, el color representa el tipo de enfermedad y la altura se refiere al número de visitas para la enfermedad correspondiente. Esto permite a los usuarios captar rápidamente información sobre las principales enfermedades del paciente, frecuencia de visitas, etc. Dado que hay muchos tipos de enfermedades posibles, para evitar el desorden visual, damos diferentes colores a las 5 enfermedades principales, mientras que el resto se representan en gris.

Los usuarios pueden ver información específica sobre enfermedades a través de la leyenda del lado derecho. El número de pacientes en el área visible se puede ajustar usando los botones más o menos a la derecha, y al hacer clic en el botón de pantalla completa se muestran todos los pacientes seleccionados en la ventana actual.

Para representar el comportamiento de covisita entre pacientes, diseñamos un enlace de covisita para mostrar explícitamente este comportamiento sospechoso. Como se muestra en la Fig. 4-(b), si hay un comportamiento de visita conjunta entre dos pacientes, extendemos una línea en la posición correspondiente de cada una de las líneas de tiempo de los dos pacientes y las vinculamos entre sí con una línea vertical. Además, dado que la línea de tiempo actual puede contener eventos de visitas agregados, el ancho de la línea vertical se utiliza para

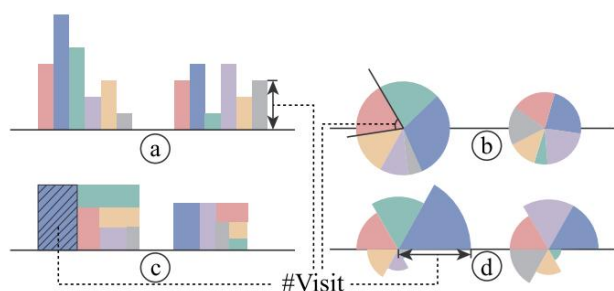


Fig. 5. Diseño alternativo del comportamiento de visita médica del paciente. El número de visitas está indicado por (a) la altura de la barra, (b) el ángulo del sector, (c) el área del rectángulo y (d) el radio del sector.

codificar el número de co-visitas. No utilizamos arcos para conectar comportamientos de co-visita porque tienden a causar más cruces y desorden visual.

Para respaldar un análisis en profundidad del comportamiento de covisita, los usuarios pueden seleccionar el umbral en el cuadro desplegable de intervalo de tiempo de covisita para cambiar a la vista de covisita (ver Fig. 6-(e)) y **examinar** todos Comportamientos de covisita dentro del período de tiempo seleccionado. Cuando el mouse pasa sobre la barra, aparece más información contextual sobre esta co-visita, incluido el tiempo de visita de cada paciente, la institución médica, la lista de medicamentos, etc. Los usuarios pueden verificar la edad, el sexo, la enfermedad y los costos de los medicamentos de un paciente haciendo clic en la identificación del paciente. Esta información de respaldo ayuda a los usuarios a razonar y verificar si están cometiendo fraude colusorio. Si la revisión está completa, los usuarios pueden hacer clic en el botón de lápiz en la esquina superior derecha para anotar a estos pacientes con un motivo en la página de etiquetado emergente.

Diseño alternativo: antes de finalizar el gráfico de barras para el registro de visitas del paciente, teníamos tres diseños alternativos: el gráfico circular, el mapa de árbol y el gráfico de rosas de Nightingale. En la Fig. 5, el número de visitas se indica por altura, ángulo, área y radio, respectivamente. Aunque las tres últimas representaciones son más compactas, los elementos visuales que representan cada enfermedad tienen diferentes ángulos, lo que dificulta las comparaciones. Además, para el gráfico circular y el gráfico de rosas de Nightingale, el radio y el área están al cuadrado, lo que puede engañar fácilmente a los usuarios. El gráfico de barras, por el contrario, tiene una orientación fija, lo que resulta adecuado para comparar enfermedades.

VI. EVALUACIÓN

Para demostrar la eficacia de FraudAuditor, realizamos dos estudios de caso y entrevistas a expertos utilizando el conjunto de datos de seguros médicos del mundo real. Tomamos una muestra de los registros de 1035 pacientes en un distrito de 2019 a 2020, que contenían más de 46 000 registros de visitas y más de 300 000 registros de medicamentos. En escenarios reales, los expertos seleccionarán datos de escala similar para analizarlos mediante filtrado espacio-temporal.

A. Estudios de caso

1) Examen de Crafty Fraud Group: un auditor de seguros de salud quería detectar patrones complejos de comportamiento fraudulento. Como se muestra en la Fig. 3-(a1), eligió la farmacia y el hospital comunitario en la vista de atributos del paciente, ya que están mal regulados y son propensos al fraude. Notó una disminución significativa en el número

de pacientes con 1-20 visitas. La mayoría visitó farmacias y hospitales comunitarios, concentrada entre 21 y 60 veces.

Con base en el conocimiento previo, estableció el intervalo de tiempo máximo para una covisita en 1 hora y el número de covisitas en al menos 2. Para evitar pasar por alto el fraude colusorio, excluyendo al mismo tiempo a pacientes individuales y grupos pequeños de solo dos pacientes, fijó el tamaño mínimo del grupo en 3 y ajustó el número mínimo de covisitas de 4 a 2 para generar la red de covisitas. Luego hizo clic en el botón para generar la red de covisitas y examinó los resultados de los modelos de detección de fraude colusorio, que contenían 48 grupos.

Como se muestra en la Fig. 6-(b), para filtrar los grupos con intervalos de visita cortos, seleccionó la barra de 0-1 minuto en el gráfico sobre la brecha mínima dentro de la co-visita en la vista de atributos del grupo.

En los grupos restantes, notó que los dos grupos de la red de visitas conjuntas de pacientes están muy conectados (Fig. 6-(a)). Cuatro puntos circundantes conectados a ellos despertaron el interés del experto y, por lo tanto, se añadieron para su análisis. En la vista de comparación de pacientes, notó que los pacientes seleccionados tenían una baja similitud de enfermedades y una alta similitud de medicamentos (Fig. 3-(c)).

Para entender por qué las enfermedades y los medicamentos no correspondían, seleccionó a todos los pacientes y descubrió que sus honorarios fluctuaban enormemente.

Luego pasó a la vista de comportamiento del paciente para investigar los detalles de las visitas. Notó un aumento en la línea de tiempo del número de visitas y la identificación del paciente de co-visitas en la visualización de la secuencia de visitas alrededor de diciembre de 2019 (Fig. 3-(d)). Los pacientes P-0037 y P-0960, en particular, no tenían registros médicos desde enero de 2019 hasta noviembre de 2019 y rara vez visitaron instituciones médicas después de 2020. Su primera visita médica fue una visita conjunta en diciembre de 2019.

Así que redujo el rango de tiempo seleccionado (Fig. 6-(d)). Después de ajustar la granularidad del tiempo a una semana, notó múltiples visitas conjuntas en las semanas del 19 y 26 de diciembre, lo que indica visitas conjuntas frecuentes y sospechosas en un período corto. Visitas tan frecuentes en tan poco tiempo resultaban muy sospechosas. Así que ajustó el intervalo de tiempo de las visitas conjuntas a 15 minutos y descubrió que todas las visitas conjuntas de estos pacientes se realizaron en la Institución 9505010 (Fig. 6-(e)), que era la farmacia que visitaban con frecuencia (Fig. 6-(c)). Pasó el cursor sobre la barra de covisitas para verificar el horario específico de las visitas y descubrió que el tiempo para comprar medicamentos estaba muy cerca, y muchas veces dentro de 1 minuto. El costo de cada visita también estuvo dentro de un rango determinado. Hizo clic en los pacientes para ver la información específica sobre enfermedades y tarifas. Los resultados mostraron que todas las enfermedades no tenían una directividad específica y eran complicadas, y el costo de cada visita era relativamente fijo. Por lo tanto, especuló que estos pacientes probablemente eran un grupo de fraude colusorio y los etiquetaron con las etiquetas correspondientes. La investigación posterior confirmó que estos estafadores se confabulaban con los vendedores para ganar dinero comprando con frecuencia medicamentos para revenderlos y al mismo tiempo ayudan a la farmacia a aumentar las ventas a cambio de sobornos.

2) Exclusión de grupos de falsos positivos con enfermedades crónicas: Otro auditor de seguros de salud quería excluir los grupos de falsos positivos, clasificados erróneamente como fraude colusorio debido al comportamiento de visitas colectivas por enfermedades crónicas. Después de revisar la vista de los atributos del paciente, desmarcó a los pacientes menores de 30 años que eran menos propensos a enfermedades crónicas. El gráfico de barras sobre el número de visitas y el número de personas mostró que la mayoría de los pacientes con más de 60 visitas eran

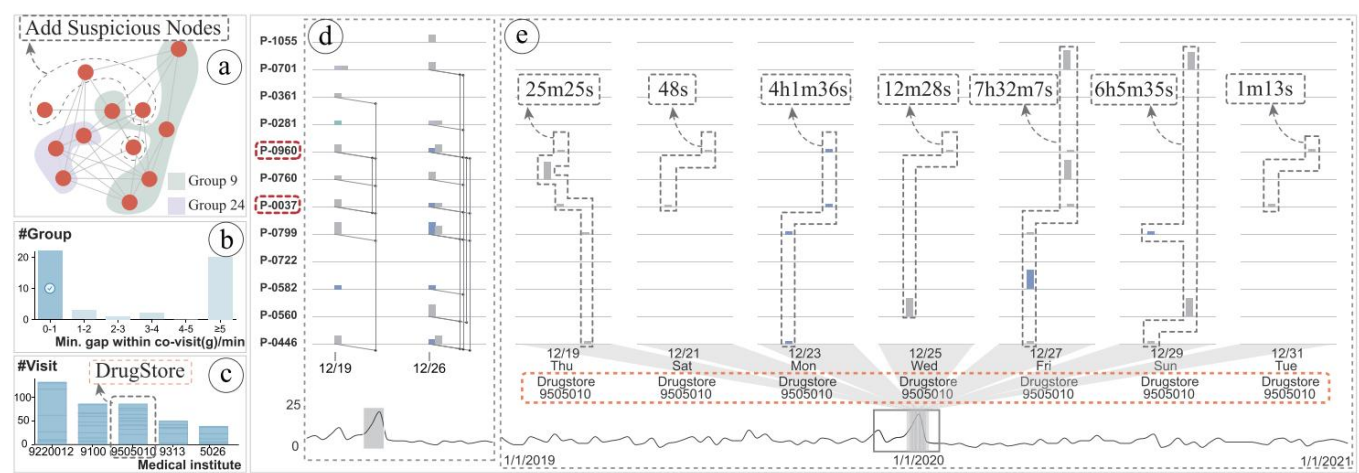


Fig. 6. El proceso de análisis visual del caso 1 consiste principalmente en (a) seleccionar dos grupos sospechosos y sus vecinos, (b) filtrar grupos, (c) explorar los atributos de los pacientes, (d) verificar registros frecuentes de visitas conjuntas de corta duración, duración, y (e) verificar los comportamientos de co-visita en múltiples períodos de tiempo.

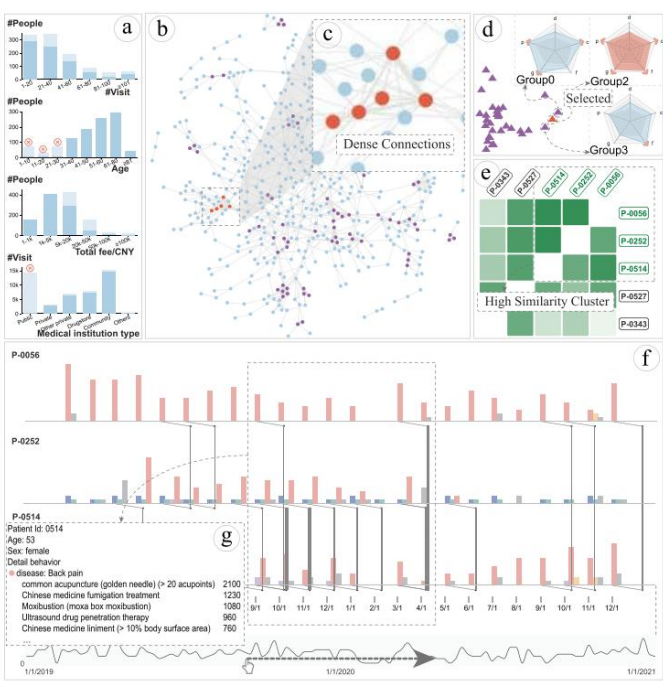


Fig. 7. El proceso de análisis visual en el caso 2 consiste principalmente en (a) filtrar pacientes según atributos, (b) detectar grupos, (c) identificar grupos según conexiones, (d) seleccionar grupos sospechosos, (e) comparar similitud intragrupo, (f) explorar comportamientos detallados de los pacientes y (g) razonamiento a través de información contextual.

Los pacientes retenidos después del filtrado y también fueron los principales visitantes de varios institutos médicos. Luego de comprender las características básicas, procedió a observar la red de visitas conjuntas entre estos pacientes. La red resultante mostró que casi todos estaban conectados. Ella atribuyó esto principalmente al hecho de que los datos contenían numerosas visitas a hospitales públicos, lo que facilitaba que los pacientes tuvieran visitas conjuntas con otras personas. Por lo tanto, excluyó los registros de visitas a hospitales públicos (aún se conservaban las visitas a otras instituciones) (Fig. 7-(a)) porque el gran número de visitas a estos hospitales fácilmente podría llevar a

a las conexiones entre pacientes y la probabilidad de fraude es menor en instituciones bien reguladas. La red regenerada tenía significativamente menos bordes (Fig. 7-(b)) y los grupos sospechosos marcados en violeta eran claramente visibles.

Luego encontró algunos valores atípicos en los resultados de la proyección grupal. Pasó el cursor sobre un valor atípico y descubrió que ocupaba el puesto número 2 en la lista del grupo y que tenía 4 métricas anómalas en el gráfico de radar (Fig. 7-(d)), como 140 visitas conjuntas. Entonces hizo clic en ese gráfico de radar y decidió analizar más a fondo a los pacientes dentro de ese grupo. Utilizando el gráfico de barras de enfermedades, descubrió que las principales enfermedades en este grupo eran enfermedades crónicas como el dolor de espalda y la hipertensión. Los cinco medicamentos más comprados incluyeron acupuntura, medicina tradicional china de fumigación y Qianghuo 6, compatible con enfermedades crónicas. En la vista matricial, después de agrupar a los pacientes por enfermedad, encontró que P0056, P0252 y P0514 tenían una alta similitud de enfermedad y fármaco (Fig. 7-(e)), y P0514 y P0056 tenían hasta un 90% de similitud de enfermedad y el resto tiene más del 70% de similitud de enfermedades entre ellos. Luego seleccionó a estos tres pacientes y el área resaltada en el gráfico de barras mostró su distribución en cada atributo en relación con todo el grupo. La mayoría de las visitas de estos pacientes se distribuyeron en un hospital comunitario número 9202. El gráfico de área sobre tarifas refleja que sus gastos de seguro médico son relativamente estables.

Sospechaba que los tres pacientes podrían haber sido detectados por error como un grupo de fraude colusorio, por lo que pasó a la vista de comportamiento del paciente para investigar el comportamiento de las visitas y obtener una mayor confirmación. Descubrió que las barras rosadas que representaban el dolor de espalda dominaban la visualización de la secuencia de visitas y aparecían a intervalos regulares (Fig. 7-(f)). Entre septiembre de 2019 y marzo de 2020, hubo múltiples vínculos de visitas conjuntas entre P0252 y P0514, por lo que pasó la línea de tiempo a este período. Desde el punto de vista de la visita conjunta, supo que estos pacientes visitaban el hospital comunitario (ID 9202) aproximadamente cada semana. La altura de la barra también mostró que gastaron aproximadamente la misma cantidad cada vez, en consonancia con el ciclo habitual de cambio de fármaco.

Al hacer clic en los pacientes P0252 y P0514, se reveló que sus edades y honorarios totales eran similares (Fig. 7-(g)), por lo que dedujo que podrían ser pacientes que se conocían y acudieron juntos al médico. Entonces marcó el grupo como normal en la vista de anotaciones y dio el motivo correspondiente.

B. Entrevista con expertos

Para evaluar la utilidad y eficacia de FraudAuditor, realizamos entrevistas con seis expertos en el campo. Estos expertos son nuestros colaboradores pero no han participado en el proceso de diseño. Dos de ellos (E1, E2) son expertos en productos relacionados con la salud y conocen los escenarios reales y el negocio de los seguros de salud. Los otros cuatro (E3-E6) son expertos en detección de fraude en seguros médicos, comprensión de patrones de fraude comunes y algoritmos de detección.

Procedimiento: Cada entrevista tuvo una duración de 90 minutos. Primero, comenzamos con una capacitación de 30 minutos sobre el propósito y el flujo de uso de FraudAuditor. Luego, los expertos dedicaron 45 minutos a explorar y descubrir sospechas de fraude grupal utilizando el sistema. En los últimos 15 minutos, le pedimos a cada experto que completara un cuestionario que consta de evaluaciones cuantitativas basadas en una escala Lik-ert de cinco puntos y preguntas cualitativas. Resumimos las ideas en los siguientes cuatro aspectos.

Modelo de minería de grupos sospechosos: los expertos coinciden en que el algoritmo de detección, considerando tanto el número de covisitas como el intervalo de tiempo de una covisita, puede identificar grupos sospechosos con mayor precisión (calificaciones: 4/5). Los expertos dijeron que la mayoría de los algoritmos automatizados existentes eran cajas negras sin explicación mecánica. Con FraudAuditor, los expertos pueden comprender los mecanismos de detección y ajustarlos de forma interactiva (por ejemplo, estableciendo el número de covisitas y el umbral de intervalo de tiempo de una covisita en función de su propia experiencia). Los efectos de los diferentes parámetros del algoritmo en los resultados son visibles en tiempo real, lo que puede aumentar la confianza de los expertos en el algoritmo. E3 dijo: "La red de covisitas puede ayudar a verificar muchas de mis hipótesis. Por ejemplo, los hospitales públicos provocan muchas covisitas y excluir registros en ellos no tendría mucho efecto en el número de grupos". E3 añadió que "si el algoritmo puede tener en cuenta la continuidad de las visitas, será más eficaz".

Visualización e interacción: los expertos creen que FraudAuditor puede satisfacer las necesidades de análisis. FraudAuditor utiliza visualizaciones básicas (p. ej., gráficos de barras, gráficos de líneas y diagramas de enlaces de nodos) y métricas estadísticas comunes (p. ej., mediana y media). Dado que los expertos en atención médica no están familiarizados con la visualización, estos diseños visuales simples reducen la barrera del aprendizaje. E1 mencionó que "la vista del comportamiento del paciente (calificaciones: 4,5/5) tiene un diseño hermoso y práctico, lo que puede ayudar a evaluar la evolución de la enfermedad y el comportamiento de las visitas conjuntas". E4 señaló que "se necesita algo de exploración para comprender la coordinación del sistema. Nunca antes había usado funciones similares y poco a poco me fui familiarizando con ellas después de un período de uso". E2 afirmó que "normalmente no tenemos mucho tiempo para verificar cada grupo en detalle en el trabajo real. El filtrado interactivo de pacientes (calificaciones: 4,3/5) y grupos en el sistema, así como la clasificación en el gráfico radar personalizado (calificaciones: 4/5), me permitió seleccionar algunos grupos de mayor sospecha, lo que mejoró mi eficiencia de análisis". La retroalimentación de la

Los expertos respaldan la utilidad de FraudAuditor para ayudar a descubrir patrones en los datos de manera eficiente sin tener que consultar e interpretar directamente los aburridos datos sin procesar.

Usabilidad del sistema: Los expertos coinciden en que el enfoque de análisis visual propuesto sigue un flujo de trabajo de auditoría de seguros de salud real. Desde descripciones generales hasta detalles, cada paso es fácil de entender (calificaciones: 4,3/5). "En comparación con el proceso de auditoría tradicional, el sistema puede acortar en gran medida el tiempo de análisis manual y la información proporcionada por el sistema es lo que necesito en el análisis", comentó E1. E6 comentó: "el sistema ofrece interacciones fáciles de usar y el proceso de análisis es razonable". También se presentaron evaluaciones sobre el costo del aprendizaje. E5 afirmó: "Cuando comencé a utilizar este sistema, tenía que leer la documentación con frecuencia. El sistema tiene una interfaz diferente al sistema de gestión de bases de datos del seguro médico. Se recomienda que el sistema de análisis visual guíe a los usuarios durante el uso". E5 añadió: "Quiero utilizar el sistema para mi trabajo. Una vez que esté familiarizado con el proceso, podré evaluar mejor la eficacia de los algoritmos de detección de fraude". Basándose en los comentarios de los expertos, FraudAuditor puede ayudar a los expertos a detectar, analizar y etiquetar grupos fraudulentos.

Sugerencias: Los expertos también ofrecieron valiosas sugerencias para FraudAuditor. E3 sugirió que el sistema podría integrar más detalles de pruebas, exámenes y procedimientos en el cuidado de la salud (actualmente no disponibles en nuestro conjunto de datos), lo que ayudaría a mejorar la precisión de la verificación del fraude. E1 también solicitó opciones adicionales de filtrado de datos granulares, como seleccionar datos de una farmacia en particular, porque a veces reciben informes de casos de fraude y luego necesitan extraer de forma independiente los datos relevantes para su revisión. E2 añadió que "si el sistema pudiera admitir la detección de fraude en tiempo real y emitir advertencias oportunas de fraude, sería más eficaz para reducir la pérdida de fondos del seguro médico".

VII. DISCUSIÓN

Esta sección analiza las ventajas y limitaciones de FraudAuditor desde las perspectivas de generalización y escalabilidad. También resumimos las lecciones aprendidas del proceso de implementación y arrojam luz sobre el trabajo futuro.

Generalizabilidad: FraudAuditor está diseñado para detectar fraudes colusorios en la atención médica, donde los pacientes a menudo visitan juntos ciertos institutos médicos en momentos similares. Al ajustar la definición de covisita, el enfoque de detección puede descubrir otros tipos de fraude sanitario. Por ejemplo, en el caso de que los estafadores trabajen de forma asincrónica, los comportamientos de covisita se pueden detectar aumentando el parámetro del intervalo de tiempo de la covisita o utilizando métodos de alineación temporal (por ejemplo, DTW [28]); en el caso de que los defraudadores se dispersen espacialmente, la restricción de ubicación se puede relajar a la misma área considerando la información geográfica de los institutos médicos.

Nuestro enfoque también se puede generalizar a otros dominios de aplicaciones donde los fraudes comparten características similares de grupo y simultaneidad, por ejemplo, fraude de comercio electrónico, detección de spam y fraude de telecomunicaciones. Tomando el spam como ejemplo, los spammers siempre utilizan botnets para enviar correos electrónicos grupales controlando múltiples cuentas de bots. Similar a la red de covisitas en

seguro médico, se puede construir una red de coenvío entre cuentas de correo electrónico considerando el intervalo de envío y el número de coenvíos.

Escalabilidad: debido a los enormes volúmenes de datos sin procesar, nuestro sistema enfrenta problemas de escalabilidad en el procesamiento y visualización de datos. Permitimos a los usuarios utilizar el filtrado de datos para centrarse en un pequeño conjunto de registros. Dado que el porcentaje de fraude es pequeño, reducir el alcance de los datos mediante la segmentación espacio-temporal y el filtrado de atributos, como eliminar registros de visitas de hospitales públicos altamente regulados, está en línea con el flujo de trabajo práctico y el principio de visión general al detalle en la visualización [29]. En el caso extremo de un grupo con muchos estafadores, la visión del comportamiento del paciente puede encontrar desorden visual y cuellos de botella en la representación, que pueden mitigarse mediante el muestreo de datos y la visualización progresiva [30].

Lecciones aprendidas: En primer lugar, se deben proporcionar vistas de múltiples niveles para la visualización de datos complejos de alta dimensión para respaldar el análisis progresivo. Presentar directamente toda la información a los usuarios aumenta la carga cognitiva. Los datos sobre seguros de salud utilizados en este artículo involucran asociaciones entre múltiples sujetos y tienen una gran cantidad de dimensiones de atributos. Dividimos las tareas de análisis y las vistas coordinadas en tres niveles (es decir, nivel de descripción general, nivel de grupo y nivel de paciente). Por lo tanto, tanto el enfoque de análisis visual como el diseño del sistema siguen el principio de descripción general hasta detalle.

En segundo lugar, la visualización intuitiva y eficaz ayuda a los usuarios a aprender rápidamente. Debido a que los usuarios de nuestro sistema son expertos con experiencia en auditoría de seguros médicos, no saben mucho sobre visualización. Los gráficos de FraudAuditor son principalmente gráficos comunes y populares, como gráficos de barras y diagramas de enlaces de nodos, que ayudan a reducir el costo de aprendizaje para los usuarios y aumentar su confianza en el sistema.

Trabajo futuro: en el futuro, para detectar otros tipos de fraude, como la colusión médico-paciente, planeamos proporcionar contextos más detallados de registros médicos mediante la construcción de una red dinámica y heterogénea de pacientes, médicos e instituciones médicas.

Otra posible dirección es reducir el costo de etiquetar manualmente el conjunto de datos aprovechando técnicas de aprendizaje activo para mejorar la eficiencia de la selección de instancias de datos. Además, la precisión de la detección de grupos se puede mejorar aún más mediante algoritmos semisupervisados. También planeamos agregar más orientación y anotaciones al sistema para mejorar aún más su usabilidad.

VIII. CONCLUSIÓN

En este artículo, propusimos un enfoque de análisis visual que respalda la identificación, examen y anotación de fraude colusorio en seguros de salud. El diseño y la implementación de FraudAuditor se basan en una estrecha colaboración con expertos en el campo. Al aprovechar tanto los algoritmos automatizados como la experiencia humana, FraudAuditor admite un análisis de fraude de varios niveles, incluida la descripción general de la red de visitas conjuntas, la identificación de grupos sospechosos y el examen de pacientes sospechosos. Un conjunto de diseños de visualización respalda la detección y exploración de grupos fraudulentos. La efectividad de nuestro enfoque y la usabilidad del sistema prototipo fueron reconocidas a través de estudios de casos y entrevistas con expertos en auditoría de seguros de salud.

REFERENCIAS

[1] J. Li, K.-Y. Huang, J. Jin y J. Shi, "Una encuesta sobre métodos estadísticos para la detección de fraude en la atención médica", *Health Care Manage. Ciencia*, vol. 11, núm. 3, págs. 275–287, 2008.

[2] L. Akoglu, M. McGlohon y C. Faloutsos, "OddBall: Detección de anomalías en gráficos ponderados", en *Proc. Conferencia Asia-Pacífico Conocimiento. Descubrimiento. Minería de datos*, 2010, págs. 410–421.

[3] P. Bindu, R. Mishra y PS Thilagam, "Descubriendo comunidades de spammers en Twitter", *J. Intell. inf. Sistema*, vol. 51, núm. 3, págs. 503–527, 2018.

[4] Z. Niu, D. Cheng, L. Zhang y J. Zhang, "Análisis visual para la gestión del riesgo de préstamos con garantía en red", en *Proc. Visual del Pacífico. Symp.*, 2018, págs. 160-169.

[5] S. Chen y A. Gangopadhyay, "Un enfoque novedoso para descubrir fraudes en la atención médica mediante análisis espectral", en *Proc. En t. Conf. Información sanitaria*, 2013, págs. 499–504.

[6] J. Wang, R. Wen, C. Wu, Y. Huang y J. Xiong, "FdGars: Detección de estafadores mediante redes convolucionales de gráficos en un sistema de revisión de aplicaciones en línea", en *Proc. Conferencia World Wide Web*, 2019, págs. 310–316.

[7] B. Xu, H. Shen, B. Sun, R. An, Q. Cao y X. Cheng, "Hacia la detección de fraude en préstamos al consumo: redes neuronales gráficas con campo aleatorio condicional restringido por roles", en *Proc. Conferencia AAAI. Artif. Intel.*, 2021, págs. 4537–4545.

[8] Q. Zhong et al., "Detección de morosos financieros en pagos de crédito en línea a través de una red de información heterogénea atribuida de múltiples vistas", en *Proc. Web Conf.*, 2020, págs. 785–795.

[9] I. Molloy et al., "Análisis de gráficos para la puntuación en tiempo real del fraude transaccional entre canales", en *Proc. En t. Conf. Criptogr. financiero. Seguridad de datos*, 2016, págs. 22–40.

[10] Z. Li, H. Xiong y Y. Liu, "Minería de patrones de volcanes y agujeros negros en gráficos dirigidos: un enfoque general", *Data Mining Knowl. Descubrimiento*, vol. 25, núm. 3, págs. 577–602, 2012.

[11] H. Joudaki et al., "Uso de la minería de datos para detectar fraude y abuso en la atención médica: una revisión de la literatura", *Glob. J. Ciencias de la Salud*, vol. 7, núm. 1, págs. 194-202, 2015.

[12] L. Akoglu, H. Tong y D. Koutra, "Detección y descripción de anomalías basadas en gráficos: una encuesta", *Data Mining Knowl. Descubrimiento*, vol. 29, núm. 3, págs. 626–688, 2015.

[13] B. Zhao, Y. Shi, K. Zhang y Z. Yan, "Detección de anomalías en seguros médicos basada en una red de información dinámica heterogénea", en *Proc. IEEE Internacional. Conf. Bioinf. Biomed.*, 2019, págs. 1118-1122.

[14] K. Ding, J. Li, R. Bhanushali y H. Liu, "Detección profunda de anomalías en redes atribuidas", en *Proc. Internacional SIAM. Conf. Minería de datos, SIAM*, 2019, págs.

[15] D. Wang et al., "Una red atenta a gráficos semisupervisados para la detección de fraude financiero", en *Proc. En t. Conf. Minería de datos*, 2019, págs. 598–607.

[16] S. Ko et al., "Análisis de enlaces de red multivariados de alta dimensión con detección, resaltado y exploración de anomalías integradas", en *Proc. Conferencia IEEE. Vis. Ciencia analítica. Technol.*, 2014, págs. 83–92.

[17] N. Cao, C. Shi, S. Lin, J. Lu, Y.-R. Lin y C.-Y. Lin, "TargetVue: análisis visual de comportamientos anómalos de usuarios en sistemas de comunicación en línea", *IEEE Trans. Visual. Computadora. Gráfico.*, vol. 22, núm. 1, págs. 280–289, enero de 2016.

[18] C. Maças, E. Polisciuc y P. Machado, "VaBank: análisis visual para transacciones bancarias", en *Proc. En t. Conf. inf. Visualización*, 2020, págs. 336–343.

[19] C. Maças, E. Polisciuc y P. Machado, "ATOVis: una herramienta de visualización para la detección de fraude financiero", *Inf. Visual.*, vol. 21, núm. 4, págs. 371–392, 2022.

[20] J. Zhao, N. Cao, Z. Wen, Y. Song, Y.-R. Lin y C. Collins, "#FluxFlow: análisis visual de información anómala que se difunde en las redes sociales". Traducción IEEE. Visual. Computadora. Gráfico., vol. 20, núm. 12, págs. 1773–1782, diciembre de 2014.

[21] E. Bertini, P. Hertzog y D. Lalanne, "SpiralView: Hacia la evaluación de políticas de seguridad a través de la correlación visual de los recursos de la red con la evolución de las alarmas", en *Proc. Síntoma IEEE. Vis. Ciencia analítica. Technol.*, 2007, págs. 139-146.

[22] P. Silva, C. Maças, E. Polisciuc y P. Machado, "Herramienta de visualización para apoyar la detección de fraude", en *Proc. En t. Conf. inf. Visualización*, 2021, págs. 77–87.

[23] Y. Lin et al., "TaxThemis: minería interactiva y exploración de grupos sospechosos de evasión fiscal", *IEEE Trans. Visual. Computadora. Gráfico.*, vol. 27, núm. 2, págs. 849–859, 2021.

[24] W. Didimo, G. Liotta, F. Montecchiani y P. Palladino, "Un sistema avanzado de visualización de redes para la detección de delitos financieros", en *Proc. Visual del Pacífico. Symp.*, 2011, págs. 203–210.

- [25] J. Tao et al., "Análisis visual de anomalías colectivas a través de un gráfico de correlación de alto orden", en Proc. Visual del Pacífico. Symp., 2018, págs. 150-159.
- [26] VD Blondel, J.-L. Guillaume, R. Lambiotte y E. Lefebvre, "Despliegue rápido de comunidades en grandes redes", J. Statist. Mecánica: Experimento teórico, vol. 2008, núm. 10 de 2008, art. No. P10008.
- [27] B. Schölkopf, A. Smola y K.-R. Müller, "Análisis de componentes no lineales como problema de valores propios del núcleo", Neural Comput., vol. 10, núm. 5, págs. 1299-1319, 1998.
- [28] M. Müller, "Dynamic time warping", Recuperación de información para música y movimiento. Berlín, Alemania: Springer, 2007, págs. 69-84.
- [29] B. Shneiderman, "Los ojos lo tienen: una tarea por taxonomía de tipos de datos para visualizaciones de información", en The Craft of Information Visualization. Amsterdam, Países Bajos: Elsevier, 2003, págs. 364-371.
- [30] E. Zraggen, A. Galakatos, A. Crotty, J.-D. Fekete y T. Kraska, "Cómo las visualizaciones progresivas afectan el análisis exploratorio", IEEE Trans. Visual. Computadora. Gráfico., vol. 23, núm. 8, págs. 1977-1987, agosto de 2017.



Zihan Zhou recibió una licenciatura en tecnología de medios digitales de la Universidad de Zhejiang en 2021. Actualmente está trabajando para obtener un doctorado en el State Key Lab de CAD&CG de la Universidad de Zhejiang, Hangzhou. Sus intereses de investigación son la visualización de datos y el análisis visual.



Dongming Han recibió su licenciatura en ingeniería de software de la Universidad de Zhejiang en 2017. Actualmente está trabajando para obtener un doctorado en el State Key Lab de CAD&CG de la Universidad de Zhejiang, Hangzhou, China. Sus intereses de investigación incluyen visualización de información, visualización de gráficos y análisis visual.



Jiehui Zhou recibió su licenciatura en informática y tecnología de la Universidad Central South. Actualmente está trabajando para obtener el título de doctorado en el Laboratorio Estatal Clave de CAD&CG de la Universidad de Zhejiang y trabaja bajo la supervisión del Prof. Wei Chen. Su principal investigación se centra en el análisis visual, la inteligencia de decisiones y la IA centrada en el ser humano, con especial interés en sus aplicaciones en el ámbito sanitario.



Xumeng Wang recibió el doctorado en informática y tecnología de la Universidad de Zhejiang en 2021. Es profesora de informática en la Universidad de Nankai. Sus intereses de investigación son el análisis visual y la preservación de la privacidad.



Haochao Ying recibió su licenciatura en informática y tecnología de la Universidad Tecnológica de Zhejiang en 2014 y su doctorado en la Facultad de Ciencias de la Computación de la Universidad de Zhejiang en 2019. Actualmente es profesor asistente en la Facultad de Salud Pública. Universidad de Zhejiang.

Sus intereses de investigación incluyen la minería de datos para el cuidado de la salud y sistemas de recomendación personalizados. Es autor de algunos artículos en prestigiosas conferencias y revistas internacionales, como IEEE Transactions on Knowledge and Data Engineering, IEEE/ACM Transactions on Computational Biology and Bioinformatics, Journal of Biomedical and Health Informatics, IJCAI, ICML y CVPR.



Jie Wang recibió una maestría en ingeniería de software de la Universidad de Zhejiang en 2021. Es ingeniero de desarrollo de inteligencia empresarial en Alibaba Group, Hangzhou. Su investigación se centra en la visualización de información y el análisis aumentado.



Jian Wu recibió su doctorado en ciencias y tecnología informática de la Universidad de Zhejiang. Es profesor titular en la Universidad de Zhejiang. Actualmente es el director del Centro de Investigación Real Doctor AI de la Universidad de Zhejiang. Sus intereses de investigación incluyen la inteligencia artificial, la minería de datos y sus aplicaciones en salud y biomedicina. Ha publicado más de 200 artículos en algunas revistas arbitradas de prestigio y actas de congresos, como IEEE Transactions on Knowledge and Data Engineering, IEEE Transactions on Medical Imaging, CVPR, IJCAI, AAAI, ICML y MICCAI. Es un miembro distinguido de la CCF.



Hui Ye recibió una maestría en ingeniería de software de la Universidad de Zhejiang en 2021. Es ingeniera de desarrollo en Tencent, Shenzhen. Su investigación se centra en la visualización de datos.



Huanliang Wang está trabajando actualmente para obtener una maestría en el State Key Lab de CAD&CG de la Universidad de Zhejiang, Hangzhou, China. Sus intereses de investigación son la visualización de datos y el análisis visual.



Wei Chen es profesor del State Key Lab de CAD&CG de la Universidad de Zhejiang. Sus intereses de investigación incluyen visualización y análisis visual. Ha publicado más de 90 artículos sobre IEEE/ACM Transactions e IEEE VIS. Se desempeñó activamente como editor invitado o asociado de IEEE Transactions on Visualization and Computer Graphics, IEEE Transactions on Intelligent Transportation Systems y Journal of Visualization. Para obtener más información, consulte <http://www.cad.zju.edu.cn/home/chenwei/>