|  |  |
| --- | --- |
|  | Detecting tax evasion: a co-evolutionary approach |

Resumen:

Presentamos un algoritmo que puede anticipar la evasión fiscal modelando la coevolución de los esquemas tributarios con políticas de auditoría. El incumplimiento impositivo malintencionado, o evasión, representa miles de millones de ingresos perdidos cada año. Desafortunadamente, cuando los administradores tributarios cambian las leyes tributarias o los procedimientos de auditoría para eliminar esquemas fraudulentos conocidos, se implementa otro esquema potencialmente más rentable. Por lo tanto, modelar los esquemas fiscales y las políticas de auditoría dentro de un marco único puede proporcionar grandes ventajas. En particular, podemos explorar las formas probables de los esquemas fiscales en respuesta a los cambios en las políticas de auditoría. Esto puede servir como un sistema de alerta temprana para ayudar a enfocar los esfuerzos de cumplimiento. Además, las políticas de auditoría pueden ajustarse para ayudar a mejorar la detección del esquema fiscal. Demostramos nuestro enfoque utilizando el esquema fiscal iBOB y demostramos que puede capturar la coevolución entre la evasión fiscal y la política de auditoría. Nuestros experimentos muestran el comportamiento oscilatorio esperado de un sistema biológico de evolución conjunta.

Introducción:

La brecha fiscal bruta de 2006 en los Estados Unidos, es decir, la diferencia entre el impuesto adeudado y el impuesto pagado a tiempo, se estimó en $ 450 mil millones (IRS 2006). La Oficina de Responsabilidad del Gobierno (GAO, por sus siglas en inglés) estima además que $ 91 mil millones de esta brecha fiscal pueden atribuirse a los ingresos ocultos en refugios fiscales compuestos por múltiples entidades de "transferencia", como sociedades, corporaciones y fideicomisos (GAO 2014a). Las empresas financieras y legales rastrean el código fiscal en busca de ambigüedades para descubrir y promover refugios fiscales abusivos. Dichas estrategias de evasión de impuestos ilegales utilizan transacciones complejas dentro de las redes de entidades fiscales que están diseñadas para reducir y ocultar las obligaciones tributarias de sus accionistas individuales. Para los fines de la discusión aquí, clasificamos tales secuencias de transacciones como 'esquemas de evasión de impuestos' en contraste con las estrategias legales de evasión de impuestos que se adhieren a la letra y la intención del código tributario.

Si bien los auditores tributarios tienen ejemplos históricos de esquemas tributarios para ayudar a guiar los esfuerzos de examen, los promotores de refugios tributarios a menudo adaptan sus estrategias a medida que se descubren los esquemas existentes y / o cuando se hacen cambios a las regulaciones tributarias existentes. Un ejemplo notable es el llamado refugio fiscal BOSS (Estrategias de venta de bonos y opciones) que se promovió ampliamente pero que finalmente fue rechazado. Si bien se implementaron cambios en la auditoría para detectar BOSS, no pudieron detectar la variante emergente "Hijo de BOSS" (Wright 2013). Esto es típico de la carrera de armamentos entre los evasores de impuestos y los auditores fiscales. Los importantes desafíos planteados para los esfuerzos de cumplimiento aquí han llevado a la acción reciente del Congreso a abordar algunas ambigüedades en la auditoría de la asociación y las reglas de ajuste (Contabilidad 2015).

Sin embargo, aún quedan desafíos importantes para los esfuerzos de cumplimiento que surgen de dos fuentes primarias. (a) la complejidad del código tributario: la ley tributaria no solo es cualitativamente difícil de analizar desde una perspectiva de lenguaje natural, sino que está bastante interconectada (Li et al. 2015; Katz y Bommarito 2014), como puede cuantificarse por el número de enlaces entre párrafos. Además, el cálculo de ciertas cantidades de impuestos puede ser complicado y bizantino. Por ejemplo, los propietarios de una sociedad necesitan ajustar valores de base separados y usar una de dos definiciones similares de "ingresos sustanciales incorporados" como se describe en el Código de Ingresos Internos (IRC) §734 y §743. (b) datos dispersos, sensibles y confusos: los datos de los informes tributarios se distribuyen y los estafadores ocultan deliberadamente sus intenciones. Estas personas suelen ofuscar sus esquemas mediante el uso de grandes jerarquías de entidades, por ejemplo, hasta 100 niveles y 100,000 socios (GAO 2014b), proporcionando la menor información posible y deteniendo los datos de los informes. Los auditores trabajan con datos fiscales agregados y dispersos, por ejemplo, en el formulario 1065 (Anexo K-1), muchos de los cuales podrían presentarse como documentos adjuntos en papel por separado.

En este documento describimos una metodología que puede ayudar a detectar estrategias para reducir la obligación tributaria al compensar las ganancias reales en una parte de una cartera al crear pérdidas de capital artificiales en otras partes, específicamente aquellas que utilizan la complicada ley de impuestos de sociedades en el Subcapítulo K del IRC. Si bien existen otras estrategias para la evasión fiscal abusiva que involucran acuerdos con partes relacionadas, precios de servicios o elementos jurisdiccionales de impuestos estatales y locales (SALT), consideramos el impuesto de sociedades para este análisis inicial. Los esquemas de impuestos aquí consisten en secuencias de transacciones entre entidades en redes de propiedad que, cuando se toman individualmente, parecen ser compatibles. Sin embargo, cuando todas las transacciones se combinan, no tienen otro propósito que mitigar ilegalmente la obligación tributaria y potencialmente se pueden etiquetar como evasión fiscal. Propiciamente, es posible no permitir la evasión fiscal por una de las muchas doctrinas contra el abuso, por ejemplo, la doctrina de la "sustancia económica", que especifica que las transacciones deben contener tanto la sustancia económica como un propósito comercial (Rostain y Regan 2014).

Los auditores tributarios generalmente usan protocolos para calificar si una entidad es lo suficientemente sospechosa como para someterse a una auditoría completa (Andreoni et al. 1998). Dicha auditoría completa es una investigación exhaustiva de la información financiera reportada para establecer la responsabilidad fiscal correcta (IRS 2014a). Nuestro objetivo es desarrollar algoritmos que puedan modelar la coevolución de los esquemas tributarios como lo representan las redes de transacciones con sus observables de auditoría correspondientes.

Nuestro enfoque para detectar la evasión fiscal nos coloca en el ámbito de la investigación de la IA y el razonamiento legal que comenzó en la década de 1970 (Buchanan y Headrick, 1970). El campo de la IA y el derecho (Sartor y Rotolo 2013; Bench-Capon et al. 2012) trata sobre la simulación de normas y su aparición (Aubert y Muller 2013; Dechesneetal.2013; Lotzmann et al.2013). El trabajo pionero en relación con los impuestos se produce en "Taxman" (McCarty, 1977) y, más recientemente, en el modelo de incumplimiento fiscal de la política de bienes raíces (Boer y Engers, 2013). El aprendizaje automático ha surgido recientemente como un medio para detectar transacciones sospechosas. Se basa en ejemplos históricos conocidos etiquetados para cumplir o no cumplir (DeBarr y Eyler-Walker 2006). El requisito de los datos etiquetados de esta manera para capacitación (Surden 2014; Ngaietal. 2011) limita el uso de métodos de detección para identificar ya Casos conocidos de abuso. Alternativamente, las técnicas no supervisadas se pueden usar para agrupar transacciones en función de su parecido común. Esto requiere una medida de similitud entre los diferentes tipos de transacciones, así como el conocimiento para clasificar un grupo como abusivo. En contraste, nuestro modelo no requiere datos de los contribuyentes (es similar a una búsqueda de minmax en un juego de ajedrez) y tiene la ventaja adicional de aliviar cualquier inquietud sobre la privacidad de los datos. El documento actual extiende específicamente los esfuerzos previos (Warner et al. 2014; Hemberg et al. 2015; Rosen et al. 2015) que describen la detección de evasión fiscal a través de la búsqueda evolutiva.

Procedemos de la siguiente manera para describir cómo nuestro método Simulación de la evasión fiscal y la Ley a través de la heurística (STEALTH, por sus siglas en inglés) puede utilizarse para anticipar esquemas de evasión fiscal abusivos. En la sec. 2, comenzamos con un antecedente de los fundamentos de la ley de impuestos de sociedades utilizados en STEALTH y la literatura relacionada con el modelo tributario. La metodología detallada de cómo se pueden utilizar las heurísticas coevolutivas para buscar y calificar los esquemas fiscales y los planes de auditoría se describe en la Sección. 3. En la sec. 4 demostramos la capacidad de STEALTH para identificar un esquema de impuestos de base artificial llamado

iBOB. Las conclusiones y el trabajo futuro se describen en la Sect. 5.

## Data science for assessing possible tax income manipulation: The case of Italy

## 

Resumen:

Este documento explora un tema fundamental del mundo real bajo una perspectiva de ciencia de datos. Se analiza específicamente si se puede observar el fraude o la manipulación en y desde las distribuciones del tamaño del impuesto a la renta del municipio, a través de su agregación de los informes fiscales de los ciudadanos. El caso de estudio se refiere a los datos oficiales obtenidos del Ministerio de Economía y Finanzas de Italia durante el período 2007-2011. Todas las regiones italianas (20) son consideradas. El enfoque de ciencia de los datos considerados se concreta en la adopción de la ley del primer dígito de Benford como herramienta cuantitativa. Se observan disparidades marcadas en varias regiones, lo que lleva a "conclusiones" inesperadas. Las regiones más buscadas no son las esperadas según la imaginación clásica sobre asuntos de sombra financiera de Italia.

Introducción:

Este documento aborda el tema relevante de identificar la existencia de anomalías en los ingresos fiscales. Nos centramos específicamente en el caso de las regiones italianas. El problema se enfrenta desde una perspectiva de la ciencia de datos, que es adecuada para el alcance del estudio. De hecho, la ciencia de datos representa hoy en día un área importante en la frontera de la investigación para procesar grandes conjuntos de datos (ver, por ejemplo, [17] y las referencias que contiene).

La relevancia del estudio radica en la evidencia de que evaluar los errores en los estados financieros es una tarea importante de los auditores, reguladores o analistas no solo en los mercados financieros, sino también en los asuntos macroeconómicos y públicos, como en los datos económicos gubernamentales. Los informes de datos precisos de los estados financieros son cruciales, incluso esenciales, para la gestión de los presupuestos públicos. Por lo tanto, es obligatorio observar si han ocurrido o están ocurriendo desestimaciones, errores, sesgos o incluso manipulaciones [60]. Por otro lado, los investigadores académicos deben proponer formas para detectar errores o anomalías. Se han propuesto muchos métodos y se han tomado medidas para crear y validar técnicas para evaluar diferentes constructos de errores [23]. Sin embargo, a pesar del progreso sustancial, en esta área de seguridad, los métodos disponibles presentan deficiencias que limitan su utilidad, a veces debido a hipótesis poco claras que subyacen al método. Lo más probable es que esto continúe para siempre, ya que es bien sabido que la imaginación de los delincuentes conduce a una manipulación más sofisticada, mientras que la reacción de los "responsables políticos" se ve afectada por los procesos legales. Sin embargo, los controles son desafiados por personas inteligentes, que carecen de la ética clásica.

Sin sugerir oprobio a todos los ciudadanos italianos debido a la supuesta evasión fiscal por parte de unos pocos, antes de debatir casos individuales, a menudo se admite que Italia es uno de los principales países que pierden la evasión fiscal (después de EE. UU. Y Brasil) a través de una clasificación del PIB (http://investorplace.com/investorpolitics/10-worst-countries-for-tax-evasion/#.Vvkqf3AR54c), oa través de la cantidad de pérdida fiscal como resultado de la economía sumergida. Por lo tanto, la manipulación de los ingresos se puede probar con razón en un caso (país), como en línea con los temas pertinentes discutidos recientemente, desde puntos de vista relacionados con la ética y el crimen organizado, por ejemplo, [4,15,16], o [46 ].

Por supuesto, no solo los ciudadanos falsifican sus datos financieros, sino también las empresas e incluso los gobiernos [2,41,43]. Por ejemplo, se han planteado preguntas sobre los datos presentados por Grecia al Eurostat para cumplir con los estrictos criterios de déficit establecidos por la Unión Europea (UE), ver [62], o sobre los datos macroeconómicos de China (Holz [37]).

Los gerentes pueden participar en una mayor o menor evasión de impuestos corporativos de lo que los accionistas preferirían [6]. Por otro lado, las empresas tienen incentivos para manipular las ganancias con el fin de convencer a los inversionistas, por ejemplo, que informen un número redondeado a un valor superior cuando tengan ganancias (es decir, USD 40 millones) y que informen un número como USD 39.95 millones, cuando tienen pérdidas, como lo discutió [65], habiendo observado tales patrones inusuales en las ganancias reportadas. Este enfoque de redondeo apunta a una "manipulación moderada" de los datos. Sin embargo, su relevancia no se considera despreciable para los inversores.

En el "nivel inferior", el de los ciudadanos, también se sabe que aparentemente pequeñas manipulaciones de redondeo pueden influir en la percepción de la calidad crediticia de los usuarios de estados financieros [34]. En otro nivel, donde el ciudadano está inmerso en una multitud y espera ser protegido por alguna sombra debido a un tramposo más grande, es interesante plantear la cuestión de si se puede ver un efecto colectivo. Esto se puede hacer a través del examen de las contribuciones del impuesto a la renta a nivel local. Este nivel contable es el núcleo de nuestra investigación e informe.

Bolton y Hand han proporcionado una revisión de los métodos estadísticos de detección de fraudes [13,14], mientras que las percepciones de los contadores con respecto a los métodos de prevención y detección de fraudes han sido discutidas recientemente [12]; vea también [68] para un resumen rápido o [39] para una discusión específica de un par de técnicas.

En este contexto, el enfoque de la ciencia de datos propuesto en el presente documento se basa en la ley de Benford.

Esta ley de Benford, originalmente para la distribución de conjuntos de datos del primer dígito (BL1), sigue una ley logarítmica:

Esta "ley" se deriva de las observaciones de [47] y luego de manera independiente de [11] de que la distribución del primer dígito está más concentrada en valores más pequeños: el dígito "1" tiene la frecuencia más alta, "9" la frecuencia más baja. En la Tabla 1, la frecuencia del primer dígito, según lo dado por BL1, se recuerda para la conveniencia del lector. A partir de entonces, las matemáticas pueden sugerir una ley empírica para la distribución de 2, 3, Ȇ dígitos. Sin embargo, dado que este último se vuelve rápidamente bastante uniforme, se vuelve difícil (pero está hecho) el uso de dígitos de tan alto nivel para probar la validez de los datos informados. Por lo tanto, concentremos nuestro objetivo a continuación en el primer dígito, es decir, en la validez (o no) de BL1 en un caso específico, que sirva de paradigma para otras investigaciones de big data.

En general, la ley de Benford tiene que ser recomendada porque tiene muchas ventajas, como no verse afectada por la invariancia de escala, y se admite que es útil cuando no hay ningún documento de respaldo para demostrar la autenticidad de las "transacciones" [67]. Hoy en día, esta llamada "ley" proporciona una base conveniente para el análisis digital de secuencias de números de naturaleza similar. Por ejemplo, un análisis basado en la ley de Benford se ha utilizado en una amplia variedad de formas para identificar los casos de robo y evasión de impuestos de los empleados [34] y también la desviación de los tipos de cambio de las rutas regulares [18] o de las tasas de Libor [1 ].

De hecho, después de [65], se puede obtener una "expectativa de manipulación" utilizando la ley [11], como se señaló más adelante en [48-50].

Desde [51], se admite que BL1 se puede usar para detectar fraudes en los datos contables que informan los ingresos individuales. La presentación y demostración de la ley Newcomb-Benford (1881–1938), como una poderosa metodología en el campo de la auditoría, se enfatizó aún más en [29,36,58,61], entre otros, y también recientemente en [21], Nigrini y Miller [5,8,22,45,57]. De hecho, BL1 también se aplica fuera del ámbito de la auditoría financiera; por ejemplo, ver [31] para imágenes forenses, [7] para anomalías en la tasa de natalidad, [59] para tasas de mortalidad materna, o en otras partes de las ciencias naturales [63], y sobre actividades religiosas [44]. Notamos que la literatura es enorme: ver [3], para aproximadamente la última década, [25], o [10], que contiene una lista bastante exhaustiva de referencias.

Limitándonos a la manipulación de datos del ámbito público del gobierno público, mencionemos, entre otros, la aplicación de la ley de Benford a saldos seleccionados en los Informes financieros anuales completos de los cincuenta estados de los Estados Unidos [35,38]; o las estadísticas macroeconómicas internacionales de “economía política de los números” [52]. Un estudio sobre el análisis de la distribución de dígitos de 134281 contratos emitidos por 20 unidades de gestión en dos estados, en Brasil, también encontró desviaciones significativas con respecto a la ley de Boston [24]; ver también [33]. La detección de fraude (y los métodos de prevención) en el sector público de Malasia se han discutido en [53].

En una escala más baja, la nuestra si se debe recordar, el uso de la ley de Benford ha permitido descubrir deficiencias en los datos informados por gobiernos locales, como municipios y estados en varios países, por ejemplo, EE. UU. O Brasil. Las distribuciones de dígitos de los estados financieros de 3 municipios, Valejo City, Orange County y Jefferson County, han demostrado tener desviaciones significativas de las esperadas en base a BL1 [35

En Italia, la recaudación de impuestos es una fuente fundamental de ingresos para los gobiernos locales, permitiendo la prestación eficiente de servicios [54]. Por otro lado, se sabe que la evasión fiscal está muy extendida en Italia ([15,30], Marino y Zizza [20,32]).

Obviamente, cualquier dificultad financiera de los municipios, como resultado de la evasión del impuesto a la renta, tiene graves repercusiones en la vida de los contribuyentes y empleados municipales [9]. Esto es molesto para la colectividad, por lo que parece importante tener una mejor vista de la calidad de los estados financieros y la rendición de cuentas en vista de la demanda (y el uso) de los fondos, por ejemplo, al regresar del gobierno italiano. Además, estas preocupaciones sobre la calidad de los datos, por una parte, y los procedimientos de auditoría mal reconocidos que se están utilizando en Italia, por otra parte, han resurgido vigorosamente después de la quiebra de varios organismos del gobierno local durante la reciente crisis financiera [54], - de hecho, para ser justos, más generalmente, en varios países industrializados.

Dentro de esta revisión de características contables específicas relevantes para nuestra investigación, podría ser finalmente interesante señalar al lector un caso italiano muy reciente y específico (por "casualidad"), es decir, la detección de anomalías en cuentas por cobrar y por pagar en universidades italianas, por [19]. Esto muestra que los niveles intermedios de las escalas de datos financieros pueden contener características intrigantes, por lo que además sugieren plantear preguntas sobre la detección de manipulaciones, a través de desviaciones de la ley de Benford, como aquí el caso de los ingresos fiscales en, por ejemplo, las regiones.

Por lo tanto, hemos considerado los valores agregados de los informes de impuesto a la renta de cada una de las 20 regiones de TI, - en un quinquenio reciente: [2007-2011] para el cual se dispone de los datos. Como lo sugiere [42], calibramos nuestro análisis con una prueba de 2.

Nuestro punto de vista es político-económico único: la confiabilidad contable de las contribuciones de los ciudadanos al PIB de TI, aunque las preguntas son numerosas. Con suerte, dentro de este marco, también se puede (i) ampliar el conocimiento del rango de aplicación de BL1, (ii) contribuir a una mejor aplicación de BL1 en la contabilidad e incluso (iii) indicar que se puede llegar a conclusiones socioeconómicas y políticas.

Metodología

La investigación que se presenta a continuación proporciona un análisis exhaustivo esencialmente a nivel regional del valor (= tamaño) de la distribución del impuesto a la renta entre las ciudades italianas, desde donde se contribuye al PIB del país. Específicamente, los datos de una región determinada se obtienen sumando los datos a nivel de municipios, para los municipios que pertenecen a la región en cuestión.

Nos referimos al Impuesto sobre la Renta Agregado (AIT, en lo sucesivo) de todos los ciudadanos que viven en cada región italiana. Sin embargo, la manipulación del ingreso por parte de los ciudadanos, si existe, se produciría en el nivel de Ingresos fiscales, por lo que la redacción específica y complicada del título de este documento.

El análisis numérico se realiza sobre la base de los datos oficiales obtenidos en el Ministerio de Economía y Finanzas (MEF) de Italia, y se refiere a cada año del quinquenio 2007-2011. En 2011, había 8092 municipios y 20 regiones, con características muy diferentes. Observe, de una vez, que esto se refiere a un gran número de gráficos y tablas. Uno puede concatenarlos, pero eso seguramente significa 20 visualizaciones de gráficos BL1, uno por región, cada uno con 5 conjuntos de puntos de datos, uno por año, como se ve a continuación. La ubicación del primer dígito se examinó utilizando un algoritmo simple hecho en casa. Las ocurrencias reales de cada punto de datos se compararon con las cantidades esperadas. Se utilizó MS Excel para organizar los datos en una forma utilizable.

El archivo de datos de Excel original enumeraba los municipios según el orden alfabético de sus nombres. Esto fue útil de manera pertinente en el estudio [45]. El objetivo del presente estudio es analizar los datos AIT en una meso escala, es decir, a nivel regional.

Así que primero aislamos y ubicamos los municipios en sus respectivas regiones. Usamos "http://www.comuni-italiani.it/nomi/index.html" para averiguar las regiones primarias de las municipalidades. La hoja de cálculo de Excel incluyó 20 columnas para acomodar cada año de datos para cada una de las veinte regiones y seis columnas para los cinco años de análisis más su promedio durante el quinquenio. Los primeros dígitos se extrajeron utilizando la función IZQUIERDA (texto, caracteres numéricos) de Excel al ingresar el número de la columna de interés en "texto" y "1" (para el primer dígito) en "números - caracteres".

La frecuencia de cada dígito, del 1 al 9, como primer dígito se obtiene haciendo uso de la función COUNTIF (rango, criterios) especificando el "rango" de celdas y el dígito de interés en "criterios". La frecuencia de cada dígito, de 1 a 9, como el primer dígito significativo, se determina así. Se pueden hacer varias pruebas para evaluar la validez de BL1. La prueba más clásica de χ 2 se usa como para otras aplicaciones BL1 [29]. Sin embargo, dentro de esta restricción, la visualización de los datos y las tablas de prueba informadas nos permiten identificar regularidades puntuales o anomalías: "conformidad aceptable", sugiere que el balance probablemente no esté sesgado y debe aceptarse sin más análisis; La "no conformidad" no garantiza que existan problemas en las cuentas subyacentes que conforman el AIT o que se haya producido un fraude, pero los resultados del análisis BL1 deben utilizarse como un indicador de que se necesita más investigación.

De hecho, si los datos no se ajustan a BL1, esto indica que los datos agregados pueden no ser representaciones verdaderas; los números pueden estar influenciados por las operaciones, sesgados debido a un error (nuestro, pero hicimos muchas verificaciones cruzadas), o pueden haber sido manipulados para engañar a algún usuario de los estados financieros. Desde el punto de vista de la ergodicidad [26–28,40,55,64,66], es interesante evaluar los datos a través de un promedio de tiempo; Esto ha sido tomado durante los 5 años relevantes para cada región. También se ha realizado una prueba de 2 en dichos promedios.

## Dynamics of tax evasion through an epidemic-like model.

Resumen:

En este trabajo, estudiamos un modelo de evasión fiscal. Consideramos una población fija dividida en tres compartimentos, a saber, contribuyentes honestos, evasores de impuestos y una tercera clase entre los dos mencionados, que llamamos susceptibles de convertirse en evasores. Las transiciones entre esos compartimentos están regidas por probabilidades, de manera similar a un modelo de propagación epidémica. Estas probabilidades modelan las interacciones sociales entre los individuos, así como la fiscalización del gobierno. Simulamos el modelo en gráficos totalmente conectados, así como en redes complejas aleatorias y sin escala. Para los casos de gráficos totalmente conectados y aleatorios, observamos que la aparición de evasores de impuestos en la población está asociada con una transición de fase de no equilibrio de absorción activa, que está ausente en las redes sin escala.

Introducción:

Los problemas socioeconómicos han sido el objetivo de varios estudios en los últimos años. 1, 2 Estos temas interdisciplinarios se tratan generalmente mediante simulaciones por computadora de modelos basados ​​en agentes, que nos permiten comprender la aparición de fenómenos colectivos en esos sistemas. Entre los problemas estudiados, uno de los grandes intereses es la dinámica de la evasión fiscal, que es interesante desde el punto de vista práctico porque la evasión fiscal sigue siendo un problema importante para los gobiernos. 3 - 5 Los economistas estudiaron modelos de evasión fiscal durante varios años, 6 - 13 y, más recientemente, los físicos también se interesaron en el tema 14 - 21 (para revisiones recientes, ver Refs. 3 y 4). La evidencia experimental proporcionada por Gachter sugiere que los contribuyentes tienden a condicionar su decisión sobre si pagar impuestos o no a la decisión de evasión fiscal de los miembros de su grupo. 6 Además, Frey y Torgler también proporcionan evidencia empírica sobre la relevancia de la cooperación condicional para la moral tributaria. 7

Teniendo en cuenta esas discusiones, Zaklan et al. Recientemente propuso un modelo que ha llamado la atención. 14 En el llamado modelo de Zaklan, la dinámica de los contribuyentes y los evasores de impuestos se analiza por medio del modelo de uso bidimensional a una temperatura T determinada. Cada agente i en la población artificial puede estar en uno de los dos estados posibles, a saber si ¼þ 1 (honesto) o si ¼ 1 (tramposo o evasor de impuestos). Una transición si! si (o un giro de giro) está controlado por la \ temperatura social "T y también depende de los estados vecinos más cercanos del agente (o giro) en el sitio i. Por lo tanto, para bajas temperaturas, ocurren pocos giros de giro y para altas temperaturas , se producen muchos giros de giro. En otras palabras, los evasores de impuestos tienen la mayor influencia para convertir a los ciudadanos honestos en evasores de impuestos si constituyen una mayoría en el vecindario respectivo. Además, se aplican algunas reglas de castigo: existe una probabilidad pa de un la auditoría a la que cada agente está sujeto en cada período y un período de tiempo k los evasores de impuestos detectados siguen siendo honestos. 14 En otro trabajo, la dinámica del modelo también fue controlada por otro modelo de dos estados, a saber, el modelo de mayoría de votos con ruido, 22 donde el ruido q desempeña el papel de la temperatura. En este caso, se encontraron resultados similares, 17 lo que sugiere que los resultados del modelo Zaklan son sólidos.

              Una extensión interesante de tales modelos es considerar que la transición de honesto a evasor no es abrupta. En este caso, se puede considerar un tercer estado que se puede llamar susceptible de indeciso. 10, 21 La presencia de dicha clase se analizó teniendo en cuenta la dinámica de los modelos de opinión de intercambio cinético, 21, 23 - 25 y considerando las mismas reglas de castigo del modelo de Zaklan. En este caso, se discutió 21 que la presencia de dicha tercera clase afecta sustancialmente la dinámica, y que el cumplimiento es alto por debajo del punto crítico (de la transición orden-desorden) de la dinámica de opinión gobernada por los intercambios cinéticos. Por otro lado, por encima del punto crítico, la evasión fiscal puede reducirse considerablemente por el mecanismo de aplicación.

En este trabajo, proponemos otro modelo basado en agentes de tres estados para analizar las dinámicas de evasión fiscal. Las transiciones entre las clases están regidas por probabilidades, de manera similar a lo que sucede en los modelos de propagación de la epidemia. 26 - 28 El mecanismo de cumplimiento se considera en las probabilidades mencionadas, así como la presión social de los contactos de un individuo determinado. También veremos que la aparición de evasores de impuestos en la población puede asociarse con una transición de fase de no equilibrio, que no se observó en los modelos físicos anteriores de evasión de impuestos, hasta donde sabemos.

## Using deep Q-learning to understand the tax evasion behavior of risk-averse firms

## 

Resumen

Diseñar políticas fiscales que sean efectivas para frenar la evasión fiscal y maximizar los ingresos estatales requiere una comprensión rigurosa del comportamiento del contribuyente. Este trabajo explora el problema de determinar la estrategia que se espera que siga una entidad fiscal con interés propio y adversa al riesgo, ya que "navega", en el contexto de un Proceso de Decisión de Markov, un entorno fiscal controlado por el gobierno que incluye auditorías aleatorias. Sanciones y amnistías fiscales ocasionales. Aunque las versiones simplificadas de este problema se han explorado previamente, el mero supuesto de aversión al riesgo (en oposición a la neutralidad del riesgo) aumenta la complejidad de encontrar la política óptima más allá del alcance de las técnicas analíticas. Aquí, obtenemos soluciones aproximadas a través de una combinación de Q-learning y avances recientes en Aprendizaje de refuerzo profundo. Al hacerlo, i) determinamos el comportamiento de evasión fiscal esperado de la entidad contribuyente, ii) calculamos el grado de aversión al riesgo de la entidad “promedio” que reciben las estimaciones empíricas de la evasión fiscal, y iii) evaluamos las políticas fiscales de muestra, en términos de ingresos esperados. Nuestro modelo puede ser útil como un banco de pruebas para la prueba "in vitro" de políticas fiscales, mientras que nuestros resultados llevan a varias recomendaciones de políticas.

Introducción:

Mientras se siguen sintiendo las réplicas de la última crisis financiera mundial, muchos gobiernos luchan por implementar políticas públicas debido a déficits presupuestarios o ingresos fiscales atrasados ​​(Bayer et al., 2015). El último problema surge como resultado de la reducción de la actividad económica, o cuando existe una fuerte sensación entre los contribuyentes de que el beneficio personal esperado de la evasión fiscal supera el beneficio social correspondiente de pagar impuestos (Alm y Beck, 1990; Bornstein y Rosenhead, 1990) . Esto, en ausencia de un sistema tributario y un mecanismo de cumplimiento adecuadamente diseñados, conduce a la evasión fiscal, un delito grave que socava el estado de los ingresos y socava el sentido de la justicia social, ya que los contribuyentes deshonestos parecen disfrutar de los mismos bienes públicos que los honestos. hacer. La "economía sumergida" resultante también tiene un fuerte impacto adverso en las calificaciones crediticias y los costos de préstamos (Markellos et al., 2016), programas de bienestar, políticas fiscales y desempleo (Fleming et al., 2000).

Por supuesto, los sistemas impositivos suelen contener varias salvaguardas para desalentar la evasión fiscal (definida aquí como la falta deliberada de declarar todos o algunos de los ingresos a las autoridades fiscales). Sin embargo, en la práctica, los sistemas impositivos son estructuras de políticas bastante complejas que son difíciles de hacer "herméticas" en términos de evasión fiscal, por razones que tienen que ver con i) ambigüedad ocasional en las regulaciones tributarias, lo que dificulta el cumplimiento y la aplicación impositiva (Andreoni et al., 1998), y ii) los comportamientos heterogéneos de las diversas entidades contribuyentes, en función de sus preferencias individuales de riesgo (Hokamp y Pickhardt, 2010).

Este documento se ocupa del desarrollo de un marco computacional riguroso que puede describir y predecir el comportamiento de los evasores de impuestos, asumiendo que están interesados ​​en sí mismos y trabajan para maximizar la utilidad de sus propios ingresos, equilibrando las ganancias potenciales de la evasión fiscal contra el riesgo. de quedar atrapado. En particular, estamos interesados ​​en i) estimar los ingresos estatales para cualquier conjunto dado de parámetros impositivos (p. Ej., Tasas impositivas y sanciones), ii) evaluar si las regulaciones fiscales específicas son útiles o no, y iii) predecir cómo los contribuyentes y el cumplimiento fiscal - Responder a los cambios de parámetros. Este último elemento está vinculado a la aversión al riesgo de los contribuyentes, cuyo conocimiento ayudaría al Estado a determinar los efectos de, por ejemplo, un aumento en las sanciones fiscales o las tasas de auditoría.

Las cuestiones planteadas anteriormente son esenciales para el Estado si se trata de saber en qué medida funcionan sus políticas tributarias o para clasificar políticas alternativas y tomar medidas para maximizar los ingresos. En este trabajo, proponemos explorarlos utilizando una combinación de redes neuronales profundas y Q-learning para determinar el comportamiento de evasión fiscal de un contribuyente adverso al riesgo (utilizaremos el término "empresa" en adelante porque estaremos interesados ​​principalmente en entidades comerciales ). Desarrollaremos y probaremos nuestro enfoque en un contexto que se basa en el trabajo de Goumagias et al. (2012) (donde solo se analizó el caso de neutralidad de riesgo) e implica un sistema tributario cercano al mundo real, con muchas de las trampas habituales, como tasas impositivas, auditorías aleatorias, sanciones y amnistías fiscales ocasionales, así como Heterogeneidad del contribuyente. Como veremos, la introducción de la aversión al riesgo en el modelo y la no linealidad resultante de la función de utilidad de la empresa se combinan con la dinámica de la empresa y llevan a un aumento significativo de la complejidad. Esto hace que el problema sea encontrar el comportamiento óptimo de la empresa más allá del alcance de los métodos analíticos y requiere que se apliquen poderosas técnicas de aproximación.

Las principales contribuciones de este trabajo son: i) el uso de técnicas de aprendizaje de refuerzo profundo para obtener soluciones computacionales para el comportamiento óptimo de la empresa basado en el modelo dinámico de Markov de Goumagias et al. (2012) y ii) un marco computacional para explorar el comportamiento esperado de las empresas que se interesan por el riesgo y que tienen intereses propios que pueden optar por participar en la evasión fiscal para maximizar su propia utilidad. Además, y por razones más prácticas, estimamos el coeficiente de aversión al riesgo de la firma o grupo de firmas "promedio" dados los datos empíricos sobre su cumplimiento tributario y evaluamos las políticas fiscales de muestra en términos de su beneficio para el Estado (o, equivalentemente, el nivel de evasión fiscal que dan lugar a). Según nuestro conocimiento, el nuestro es el primer trabajo para aplicar el aprendizaje profundo en el contexto de la tributación y la evasión fiscal, y el primero en obtener soluciones que revelan el comportamiento de una empresa adversa al riesgo en una escala de tiempo "fina", es decir, en un año sobre la base de su estado evolutivo a los "ojos" de las autoridades fiscales. Consideramos que nuestro enfoque es particularmente relevante tanto por el creciente interés en las aplicaciones de aprendizaje profundo como por las oportunidades que nuestro modelo brinda a los reguladores en el diseño de políticas efectivas que hacen que las entidades se comporten de manera más honesta.

El resto de este trabajo se estructura de la siguiente manera. En la Sección 2 revisamos la literatura relevante y discutimos cómo se ubica nuestro enfoque en relación con el trabajo anterior. La Sección 3 comienza con una breve descripción del sistema tributario en el que opera la empresa y explica los parámetros principales. En la misma Sección describimos un modelo basado en Markov de la evolución de la empresa a través del sistema tributario y planteamos el principal problema de optimización que nos interesa resolver y los desafíos computacionales involucrados. Nuestro enfoque de solución, que combina Q-learning y Deep Neural Networks, se detalla en la Sección 4. Finalmente, en la Sección 5 se analizan los resultados que obtuvimos, utilizando el sistema tributario griego como un estudio de caso para concretar, y su relevancia para las preguntas. planteado anteriormente con respecto al comportamiento esperado de la empresa, los incentivos para reportar ganancias, el grado de aversión al riesgo y las implicaciones políticas.

El trabajo anterior relacionado con la tributación óptima y el modelo de evasión de impuestos se puede agrupar en dos categorías principales: i) analítico (macroeconómico y basado en el principio del agente), y ii) computacional (basado en el agente, basado en la simulación).

## Tax payment default prediction using genetic algorithm-based variable selection

Resumen:

Según las estadísticas de las autoridades fiscales finlandesas, aproximadamente el 12% de todas las empresas activas en Finlandia tenían impuestos no pagados al final del año 2015. En términos monetarios, esto se traduce en más de 3 mil millones de euros en impuestos no pagados. Esta es una cantidad altamente significativa ya que el monto total de los impuestos recaudados durante 2015 fue de 49 mil millones de euros. Teniendo en cuenta la importancia económica de los impuestos impagos, se ha realizado relativamente poca investigación para identificar a las empresas que no pagan impuestos. El objetivo de este estudio es desarrollar una herramienta de apoyo a la decisión basada en algoritmos genéticos para predecir los incumplimientos de pago de impuestos. Más estrechamente, se utiliza un algoritmo genético para determinar un subconjunto óptimo o casi óptimo de variables para un modelo de análisis discriminante lineal (LDA) que clasifica a las empresas examinadas como predeterminadas o no predeterminadas. La herramienta también proporciona información acerca de la importancia de varias variables para predecir un impuesto predeterminado. El conjunto de datos consta de firmas finlandesas de responsabilidad limitada que han incumplido los impuestos de contribución al empleador o impuestos al valor agregado y el número total de variables disponibles es 72. Los resultados muestran que las variables que miden la solvencia, la liquidez y el período de pago de las cuentas por pagar comerciales son variables importantes en la predicción de impagos de impuestos. El modelo de mejor desempeño comprende tres variables transformadas no linealmente y tiene una precisión predictiva de 73.8%.

Introducción

Según las estadísticas de las autoridades fiscales finlandesas, aproximadamente el 12% de todas las empresas activas en Finlandia tenían impuestos no pagados al final del año 2015. En términos monetarios, esto se traduce en más de 3 mil millones de euros en impuestos no pagados. Esta es una cantidad altamente significativa ya que el monto total de los impuestos recaudados durante 2015 fue de 49 mil millones de euros. Además, se estima que solo se recuperará el 20% de estos impuestos impagos. Teniendo en cuenta la importancia económica de los impuestos no pagados, se ha realizado relativamente poca investigación sobre la identificación de empresas que no pagan impuestos. Sin duda, tanto las autoridades fiscales como otras partes interesadas se beneficiarían de poder predecir los incumplimientos impositivos. Por ejemplo, las autoridades fiscales podrían usar tales herramientas al seleccionar firmas para auditorías fiscales o los acreedores podrían hacer evaluaciones de riesgo basadas en la probabilidad de impagos tributarios.Si bien se ha prestado relativamente poca atención al desarrollo de modelos para la predicción de imposición de impuestos, la predicción de dificultades financieras en general ha sido un tema de investigación popular durante décadas. En un entorno finlandés, las deudas tributarias se pueden cobrar directamente y la información sobre las deudas impagas está disponible públicamente. Por lo tanto, se puede argumentar que el impago de impuestos es un fuerte indicio de dificultades financieras. Por lo tanto, los problemas de modelado en la predicción de dificultades financieras pueden, en gran medidaSe puede argumentar que el impago de impuestos es un fuerte indicador de dificultades financieras. Por lo tanto, los problemas de modelado en la predicción de dificultades financieras pueden, en gran medidaSe puede argumentar que el impago de impuestos es un fuerte indicador de dificultades financieras. Por lo tanto, los problemas de modelado en la predicción de dificultades financieras pueden, en gran medida extensión, se aplicará a la predicción de impago fiscal. Se ha utilizado una gran cantidad de variables diferentes en varios modelos de predicción de dificultades financieras, pero no hay consenso sobre qué variables son las más adecuadas para la tarea (Balcaen y Ooghe, 2006). El subconjunto de variables utilizadas en los modelos se puede seleccionar tanto por consideraciones teóricas como por resultados empíricos. El inconveniente con La selección de variables basadas en la teoría es el marco teórico limitado (Dimitras, Zanakis y Zopounidis, 1996; Lensberg, Eilifsen y McKee, 2006), mientras que la selección de variables basada en resultados empíricos puede sufrir deficiencias relacionadas con diversas cuestiones estadísticas, como la multicolinearidad ( Gilbert, Menon, & Schwartz, 1990). En línea con esto, Du Jardin (2010) mostró que hay una mejora significativa en los modelos de predicción de fallas cuando se diseñan utilizando técnicas de selección de variables apropiadas en lugar de depender de métodos comunes de la literatura financiera.

La mejor manera de determinar el subconjunto óptimo de variables para los modelos predictivos es realizar una búsqueda exhaustiva de diferentes combinaciones de variables. Sin embargo, esto a menudo no es factible ya que el número de subconjuntos crece exponencialmente con el número de variables disponibles. Los algoritmos genéticos (GA) (Holland, 1975) son un método eficiente para resolver varios problemas complejos de optimización y se han utilizado con frecuencia para la selección de características o variables en el contexto de la determinación de la salud financiera de las empresas (por ejemplo, Back, Laitinen, & Sere, 1996; Brabazon & Keenan, 2004; Gordini, 2014; Oreski & Oreski, 2014; Ravisankar, Ravi, & Bose, 2010). El objetivo de este estudio es desarrollar una herramienta de apoyo a la decisión basada en GA para predecir los incumplimientos de pago de impuestos. Más cerca,una GA se utiliza para determinar un subconjunto óptimo o casi óptimo de variables para un modelo de análisis discriminante lineal (LDA) que clasifica a las empresas examinadas como predeterminadas o no predeterminadas. Aunque un modelo de LDA tiene sus inconvenientes, investigaciones anteriores muestran que los modelos más sofisticados generalmente producen mejoras marginales más bien pequeñas (Balcaen y Ooghe, 2006; Hand, 2004). Además, un modelo LDA es relativamente fácil de usar e interpretar. Por lo tanto, este estudio se centra en un modelo LDA en lugar de enfoques más avanzados. Además del modelo LDA, el GA también proporciona una lista de frecuencia de ocurrencia de todas las variables utilizadas para desarrollar el modeloinvestigaciones anteriores muestran que los modelos más sofisticados suelen producir mejoras marginales más bien pequeñas (Balcaen y Ooghe, 2006; Hand, 2004). Además, un modelo LDA es relativamente fácil de usar e interpretar. Por lo tanto, este estudio se centra en un modelo LDA en lugar de enfoques más avanzados. Además del modelo LDA, el GA también proporciona una lista de frecuencia de ocurrencia de todas las variables utilizadas para desarrollar el modeloinvestigaciones anteriores muestran que los modelos más sofisticados suelen producir mejoras marginales más bien pequeñas (Balcaen y Ooghe, 2006; Hand, 2004). Además, un modelo LDA es relativamente fácil de usar e interpretar. Por lo tanto, este estudio se centra en un modelo LDA en lugar de enfoques más avanzados. Además del modelo LDA, el GA también proporciona una lista de frecuencia de ocurrencia de todas las variables utilizadas para desarrollar el modelo .

El resto de este estudio está organizado de la siguiente manera. En la Sección 2 se trata la literatura relacionada con los impuestos predeterminados, la predicción de dificultades financieras con los datos de los estados financieros, los algoritmos genéticos y la selección de variables. La descripción del conjunto de datos, el diseño de la investigación y la estrategia de análisis se presentan en la Sección 3 y los resultados del estudio en la Sección 4. La sección 5 concluye el estudio .

## Detecting corporate tax evasion using a hybrid intelligent system: A case study of Iran

Resumen

Este documento se concentra en la efectividad de usar un sistema inteligente híbrido que combina la red neuronal de perceptrón multicapa (MLP), la máquina de vectores de soporte (SVM) y los modelos de clasificación de regresión logística (LR) con un algoritmo de optimización de búsqueda de armonía (HS) para detectar corporativos Evasión fiscal para la Administración Tributaria Nacional de Irán (INTA). En esta investigación, la función del algoritmo de optimización es buscar y encontrar los parámetros óptimos del modelo de clasificación y la combinación de variables financieras. Nuestro sistema propuesto encuentra la estructura óptima del modelo de clasificación en función de las características del conjunto de datos importado. Este sistema se ha probado con los datos de los sectores de alimentos y textiles utilizando una estructura iterativa de 10 veces la validación cruzada que involucra a 2451 y 2053 muestras de conjuntos de prueba de las declaraciones de impuestos de un período de dos años y 1118 y 906 muestras como resultado. de la muestra utilizando las declaraciones de impuestos del año siguiente. Los resultados de los datos fuera de la muestra muestran que la red neuronal MLP en combinación con el algoritmo de optimización HS supera a otras combinaciones con 90.07% y 82.45% de precisión, 85.48% y 84.85% de sensibilidad, y 90.34% y 82.26% de especificidad, respectivamente Los sectores alimentario y textil. Además, también hay una diferencia entre los modelos seleccionados y las precisiones obtenidas en base a los datos de prueba y los datos de fuera de muestra en ambos sectores y las variables financieras seleccionadas de cada sector.

Introducción

Las declaraciones de impuestos contienen información útil para detectar la evasión fiscal. Sin embargo, gran parte de esta información es voluminosa y compleja. Por lo tanto, es necesario utilizar modelos matemáticos y estadísticos para analizar esta información. Debido a los costos asociados con la auditoría de las declaraciones de impuestos de compañías individuales, es importante encontrar modelos que identifiquen con precisión las declaraciones erróneas. En consecuencia, el objetivo de este análisis de datos es ayudar a los auditores tributarios y las autoridades tributarias a detectar compañías con una alta probabilidad de actividades declaradas erróneas para brindar a estas compañías sospechosas una auditoría más detallada. Al usar estos modelos, las autoridades fiscales pueden aumentar significativamente los ingresos fiscales y disminuir los costos de recursos humanos asociados con la auditoría manual de las declaraciones de impuestos.

Debido a los conjuntos de datos fiscales voluminosos y complejos, las autoridades fiscales de muchos países han comenzado a adoptar nuevos sistemas basados ​​en minería de datos (DM) e inteligencia artificial (AI) para ayudarles a detectar errores en los estados financieros. Debido a esto, la investigación académica en esta área (informes de investigación especialmente publicados para sistemas de detección de evasión de impuestos) se está quedando atrás y la necesidad de más atención a este tema por parte de la academia se siente especialmente más que nunca. En general, la detección de evasión de impuestos, utilizada por las autoridades fiscales para detectar declaraciones de impuestos sospechosas de evasiones fiscales, es algo similar a la detección de fraude en los estados financieros.

Las administraciones tributarias tienen que lidiar con una variedad de riesgos, como el riesgo de incumplimiento, el riesgo de evasión fiscal y el riesgo de insolvencia por parte del contribuyente. La gestión del riesgo de cumplimiento nos permite lidiar con estos riesgos observando el comportamiento de los contribuyentes. Por estas razones, los modelos predictivos inteligentes para identificar las declaraciones de impuestos para un examen adicional se pueden utilizar como una herramienta para aumentar la precisión y la eficiencia de la auditoría.

La Administración Nacional de Impuestos de Irán (INTA) es una institución gubernamental y se establece bajo la supervisión del Ministerio de Asuntos Económicos y Finanzas de Irán. El INTA brinda apoyo total para administrar los planes de reforma tributaria y ejecutar los procedimientos legales de recaudación de impuestos de manera eficiente. En el futuro, sus responsabilidades también se extenderán al monitoreo de la aplicación de la ley tributaria y las regulaciones y la creación de una base adecuada para lograr los objetivos tributarios. En este documento, presentamos un nuevo sistema inteligente híbrido para detectar la evasión de impuestos corporativos en INTA.

Nuestro objetivo es comparar la efectividad del uso de un sistema que combina la red neuronal de perceptrón multicapa (MLP), la máquina de vectores de soporte (SVM) y los modelos de clasificación de regresión logística (LR) con el algoritmo de optimización de búsqueda de armonía (HS) para detectar el impuesto corporativo Evasión por separado en dos sectores. El algoritmo de optimización de HS optimiza la selección de variables financieras además de los parámetros del modelo; por lo tanto, se puede desarrollar un modelo optimizado con una precisión aceptable en términos de una combinación de variables financieras óptimas y parámetros del modelo. El modelo híbrido fue desarrollado y probado con datos de los sectores de alimentos y textiles.

Este documento se encuentra entre las primeras investigaciones basadas en datos reales con una gran muestra de declaraciones de impuestos de diferentes sectores que operan en Irán. Para obtener un buen nivel de confianza en nuestros resultados, la salida de nuestro modelo se evaluó comparándola con los datos fuera de la muestra (además de los datos de prueba y los datos totales). Además, la validación cruzada se repitió 8 veces utilizando computación paralela para mejorar la confiabilidad de los resultados. La mayoría de los artículos publicados en esta o en áreas relacionadas se centran principalmente en resultados binarios en lugar de resultados probabilísticos. Sin embargo, los resultados binarios no son realmente útiles en la práctica. Por lo tanto, todos los modelos de clasificación presentados en este sistema propuesto generan la probabilidad de membresía en cada una de las dos clases y, a los fines de los informes, los productos se clasifican según las probabilidades obtenidas.

Esto permite que las auditorías de las empresas con mayores probabilidades se realicen con más escrutinio. Además, la comparación de las variables financieras seleccionadas en base a una cantidad de repeticiones en 10 ejecuciones del sistema para los sectores de alimentos y textiles es útil para responder la pregunta de si los patrones de comportamiento de la evasión fiscal son diferentes entre los distintos sectores.

Este documento está estructurado de la siguiente manera: los primeros sistemas inteligentes para la detección de fraudes se presentan en la Sección 2, luego los algoritmos de detección de evasión de impuestos que se usaron se describen en la Sección 3, la estructura del sistema inteligente híbrido para la detección de evasión de impuestos corporativos se explica a continuación en la Sección 4 , el método se presenta en la Sección 5, seguido de los resultados y la discusión en la Sección 6 y, finalmente, la Sección 7 proporciona conclusiones y futuras direcciones de investigación de este trabajo.

## Predictive Analytics as a Service on Tax Evasion using Gaussian Regression Process

Resumen

El análisis predictivo combina las capacidades de análisis estadístico, aprendizaje automático y extracción de datos. La gran cantidad de datos no estructurados producidos por diversos sectores públicos y privados, tales como el gobierno, los seguros de salud, las redes sociales y los académicos, permitieron que el análisis de textos permitiera comprender mejor los riesgos. El análisis predictivo puede predecir tendencias, determinar probabilidades estadísticas y actuar ante amenazas de fraude y seguridad para aplicaciones de big data como comercio de negocios, detección de fraudes, investigación de delitos, banca, seguros, seguridad empresarial, gobierno, atención médica, comercio electrónico y telecomunicaciones. El marco de análisis predictivo como servicio (PAaaS) se propone en nuestros trabajos anteriores. Este documento da solución a una de las aplicaciones de detección de fraude en los datos del impuesto sobre la renta. La solución se basa en un modelo conjunto que utiliza un proceso gaussiano con diversos parámetros hiperactivos. Las medidas de rendimiento NRMSE y COD se utilizan para analizar el modelo. Los resultados de las pruebas demostraron que los valores del tercer parámetro hiperactivo dieron un buen resultado con menos tasa de error y más varianza, lo cual es confiable para un modelo predictivo.

introducción

Los análisis de datos descubren patrones ocultos y correlaciones de grandes volúmenes de datos mediante el uso de técnicas de estadística, aprendizaje automático, inteligencia artificial y minería de datos. El análisis predictivo (PA) se refiere a predicciones como la gestión de las relaciones con los clientes, la venta cruzada, el seguro de salud, la gestión de riesgos en Banca, telecomunicaciones, detección de fraudes sobre el futuro a través de un análisis de datos. Los análisis de datos se pueden clasificar en modelos descriptivos, predictivos y prescriptivos. Los modelos descriptivos utilizan la agregación de datos para concluir sobre lo que ha sucedido, como mencionar la relación entre los datos y describe analíticas predictivas pasadas que usan modelos estadísticos para predecir el futuro como lo que puede suceder, y las analíticas prescriptivas utilizan técnicas de optimización para sugerir las formas de resultados y sus posibles efectos. Otros modelos predictivos se pueden clasificar en modelos de clasificación y modelos de regresión. Los modelos de clasificación determinan las etiquetas de clase (categóricas) donde los modelos de regresión ayudan a predecir un valor numérico. Se han desarrollado muchas técnicas para el modelado predictivo, como SVM, métodos bayesianos,Redes neuronales, modelos de regresión, k NN, modelos de elevación y árboles de decisión. Pero los modelos de conjunto demostraron tener una buena precisión en comparación con otros, por lo que entrenan varios modelos similares y combinan resultados para poder obtener un mejor modo l para predecir nuevos datos. Como se explica en [1], los modelos predictivos pueden encontrar la relación entre el resultado y las variables dependientes. De manera similar, los modelos descriptivos se utilizan para formar grupos de objetos con características similares. Hay seis fases para el proceso analítico predictivo. En la fase inicial, el proyecto se define con los resultados, los objetivos, el alcance y los resultados del proyecto. En la siguiente fase se recopilan datos de varias fuentes y se analizan. Este análisis requiere estrategias para el preprocesamiento, como la limpieza de datos,Transformación y modelado de datos para que se extraigan datos útiles para su posterior procesamiento. Posteriormente validar la hipótesis inicial utilizando modelos estadísticos. La siguiente fase es el modelo predictivo para pronosticar el futuro. Los resultados después de la implementación se pueden implementar para usarlos en la toma de decisiones del día a día. La última fase es monitorear el modelo para asegurar que está proporcionando los resultados esperados.

La detección del fraude es necesaria para cualquier sistema financiero. Hoy en día, algunas personas están engañando al gobierno al no pagar los impuestos correctamente. Una gran pérdida está siendo reportada por el gobierno. El gobierno no tiene una estimación adecuada de cuánto impuesto se debe recaudar de la gente. Si tiene una observación adecuada de los impuestos recibidos de años anteriores, el gobierno puede tomar decisiones sobre la cantidad de impuestos que se deben cobrar. Una estimación adecuada de estos montos puede facilitar la tarea de detección de fraude. Si podemos reducir el fraude asociado con la recaudación de impuestos, habrá un aumento en los ingresos para el gobierno, que se puede utilizar para actividades de desarrollo. Por lo tanto, el fraude puede detectarse extrayendo los conocimientos de los datos que se analizan en este documento.Este documento se enfoca en la capa de computación de nuestro marco de trabajo descrito en [2], que aplica el proceso gaussiano del algoritmo analítico predictivo en el conjunto de datos del impuesto a la renta para identificar el fraude en los valores tributarios proyectados. El documento está organizado de la siguiente manera: la sección 2 presenta una encuesta bibliográfica sobre varios algoritmos analíticos predictivos y diversos servicios web existentes. La sección 3 describe el marco propuesto. Las conclusiones y trabajos futuros se dan en la sección 4.

## Mining Suspicious Tax Evasion Groups in Big Data

Resumen

Existe evidencia de que un número creciente de empresas se unen para evadir impuestos de una manera no percibida. Al mismo tiempo, los datos relacionados con la información tributaria son un tipo clásico de big data. Los problemas cuestionan la efectividad de los métodos tradicionales de detección de evasión de impuestos basados ​​en la minería de datos. Para abordar este problema, primero investigamos los casos clásicos de evasión de impuestos y empleamos un método basado en gráficos para caracterizar su propiedad que describe dos rastros de relaciones sospechosas con un mismo nodo de antecedentes detrás de una Transacción de Interés Afiliado (IAT). A continuación, proponemos un modelo de red de color (CNBM) para caracterizar los comportamientos económicos, las relaciones sociales y los IAT entre los contribuyentes, y generar una Red de Intereses Interesados ​​con los Contribuyentes (TPIIN).Para llevar a cabo la tarea de detección de evasión de impuestos al descubrir grupos sospechosos en un TPIIN, se presentan métodos para construir un árbol de patrones y patrones de componentes coincidentes y se presenta la integridad de los métodos basados ​​en la teoría de grafos. Luego, describimos un experimento basado en datos reales y una red simulada. Los resultados experimentales muestran que nuestro método propuesto mejora en gran medida la eficiencia de la detección de la evasión fiscal, así como proporciona una explicación clara de los comportamientos de evasión fiscal de los grupos de contribuyentes.Los resultados experimentales muestran que nuestro método propuesto mejora en gran medida la eficiencia de la detección de la evasión fiscal, así como proporciona una explicación clara de los comportamientos de evasión fiscal de los grupos de contribuyentes.Los resultados experimentales muestran que nuestro método propuesto mejora en gran medida la eficiencia de la detección de la evasión fiscal, así como proporciona una explicación clara de los comportamientos de evasión fiscal de los grupos de contribuyentes.

introducción

La recaudación de ingresos fiscales se considera una prioridad máxima en todas las jurisdicciones nacionales y regionales [4], [10], [16], [17], [18], [19], y China no es diferente. El gobierno chino informó que la tasa de pérdida de ingresos fiscales en China es superior al 22%. Excluyendo cualquier defecto del sistema en los mecanismos de recaudación y administración de impuestos chinos, más del 12% de los ingresos fiscales en China se pierde debido a problemas técnicos.

El Sistema de Información de la Administración Tributaria de China (CTAIS, por sus siglas en inglés) se desarrolló en 1996 y desde el año 2000 está en funcionamiento, impulsando a nivel nacional una revolución en la administración tributaria apoyada por la informatización y el intercambio de datos entre diferentes provincias. El intercambio de datos ha preparado el terreno para la minería profunda y el análisis de los datos fiscales. Al mismo tiempo, China adoptó tres formas de inspección de impuestos, selección manual de casos [23], selección de casos basada en computadora (métodos basados ​​en extracción de datos [24], [7]) y selección basada en denuncias. Administración tributaria en su operación diaria de inspección tributaria. Como resultado del uso de estos métodos, los comportamientos tradicionales de evasión de impuestos, tales como escribir facturas de impuestos falsos (IVA), facturas falsas y la manipulación de cuentas se restringieron más que nunca.y el número de casos de evasión fiscal ha disminuido dramáticamente.

A lo largo de muchos años de seguimiento y análisis de datos tanto de transacciones nacionales dentro de China como de sus transacciones transfronterizas, se ha demostrado que hay una nueva tendencia a que las empresas conspiren juntas para evadir impuestos de manera indetectable s [18], [16], [4], [6], especialmente a través de transacciones legales. Llamamos a este tipo de transacciones, transacciones con intereses asociados (IAT). En el campo de la contabilidad y la gestión, se denominan "transacciones controladas". Entre las partes de la transacción, lo más importante es que existe una relación interactiva compleja y encubierta. Por ejemplo, si existe s: (A) un parentesco entre los ejecutivos o gerentes de las compañías o entre personas jurídicas o (B) una relación de interbloqueo de acciones entre los accionistas. Estas relaciones de relaciones no son solo heterogéneas,pero también diversificado y utilizado para lograr la transferencia de intereses entre empresas para evadir impuestos legítimos.

El sistema nacional de recopilación de información tributaria (NTICS, por sus siglas en inglés) desplegado en China tiene un alto volumen de transacciones y los datos relacionados. Por ejemplo, hay hay más de 31, 910,000 contribuyentes y 48,000 oficinas de administración tributaria en todo el país. El número de registros comerciales relacionados con impuestos anuales es de hasta 1.000 millones, el pico diario de estos registros es de hasta diez millones y el volumen de datos anuales agregados es de 12 TB, que se confirma a sí mismo como datos importantes. Este volumen de datos desafía los métodos tradicionales basados ​​en minería de datos para detectar la evasión fiscal. Las razones de esto son que: (A) los datos de entrenamiento deben etiquetarse manualmente, (B) los modelos entrenados generalmente son sensibles a los datos de entrenamiento, (C) los resultados de los métodos basados ​​en agrupación y en red neuronal son no explicables y no son intuitivos, (D) su eficiencia es baja ya que necesitan identificar las transacciones (incluida su información detallada) una por una,(E) el problema más importante es que algunas de las relaciones encubiertas no se registran en el NTICS, como las relaciones entre directores, gerentes y personas jurídicas. Al mismo tiempo, algunas compañías ocultan y no informan estas relaciones o demoran revelar cambios en los detalles de tales relaciones.

De acuerdo con la idea presentada por Wu et al. [23], creemos que las autoridades tributarias están equipadas con recursos limitados, y los métodos / estrategias tradicionales de auditoría tributaria consumen tiempo y son tediosos. En consecuencia, existe una necesidad apremiante de tener más aportaciones a una base de datos de evasión fiscal y recursos de información adicionales de big data (por ejemplo, relaciones con directores de empresas y parentesco entre personas jurídicas y / o directores, etc.). Esto proporcionará una base de análisis de impuestos mejorada con una técnica más razonable y permitirá el desarrollo de nuevos métodos para tratar cualquier característica nueva de los IAT, es decir, adquirir relaciones más ocultas detrás de cada IAT, mejorar la eficiencia de identificación y respaldar una representación explicable e intuitiva de los resultados minados.

Después de investigar los casos clásicos de evasión de impuestos y emplear la red de información heterogénea [25], [27], [28] para analizar sus propiedades, proponemos un modelo basado en redes de color (CNBM) para caracterizar el comportamiento económico, las relaciones sociales y las IAT entre contribuyentes. Luego, tratamos la detección de la evasión fiscal como un proceso de dos fases. El objetivo de la primera fase es descubrir los grupos sospechosos de la red de información heterogénea construida en base a la CNBM, para identificar las relaciones comerciales sospechosas. Llamamos a la primera fase como minería grupos sospechosos, MSG -fase para abreviar. En la segunda fase, los métodos tradicionales se pueden usar en todas las transacciones relacionadas con las relaciones comerciales sospechosas para detectar la evasión fiscal dentro del conjunto de grupos sospechosos. Llamamos a la segunda fase como identificación de la evasión fiscal,Fase ITE para abreviar. Los desafíos en la MSG -fase del método propuesto son: (1) cómo modelar una red heterogénea que incorpora todas las relaciones encubiertas, así como mantenerla lo suficientemente simple para ser comprensible, (2) es obvio que la diversidad de tipos de estas relaciones encubiertas no solo dan como resultado la complejidad del modelado sino que también presentan desafíos para detectar grupos sospechosos. La forma directa de representar los diversos vínculos entre los participantes en las transacciones financieras es usar colores separados para cada tipo de participante y colores separados para los distintos tipos de vínculos. En esta red heterogénea, los grupos sospechosos de evasión de impuestos aparecerán como una variedad de patrones de subgrafo y detectar estos patrones de subgraph es una clase de tareas de listado de subgraph. Sin embargo,Las diferentes formas de los grupos de evasión de impuestos sospechosos dan lugar a diferentes patrones de subgrafo, como triángulo, cuadrilátero, pentágono y hexágono, y la diferencia de color de los bordes en una patrón subgráfico específico. Por lo tanto, la detección de estos patrones subgráficos conduce a un problema de explosión combinatoria y aumenta el gasto de cálculo. Este documento intenta abordar las dificultades mencionadas anteriormente en la fase MSG del método propuesto. El MSG -phase construye el CNBM, denominado Red de Intereses Interesados ​​con los Contribuyentes (TPIIN), basado en una cierta simplificación de las relaciones y operaciones de contracción en tipos específicos de bordes de fuentes de datos. Después de construir un TPIIN, el algoritmo s para construir un árbol de patrones de la base de datos TPIIN,la generación de una base de patrones de componentes y la detección de grupos de evasión de impuestos sospechosos basados ​​en una regla para encontrar dos nodos vinculados a transacciones con el mismo antecedente se proponen para superar el problema de buscar una variedad de patrones subgráficos (el problema puede resultar en una combinación explosión). Para evaluar la efectividad del método propuesto, se realizan experimentos basados ​​en datos reales con una relación comercial adicional representada por una red simulada. Los resultados muestran que nuestro método propuesto puede mejorar en gran medida la eficiencia de la detección de evasión fiscal potencial, así como proporcionar explicaciones claras de posibles comportamientos de evasión fiscal.Los resultados muestran que nuestro método propuesto puede mejorar en gran medida la eficiencia de la detección de evasión fiscal potencial, así como proporcionar explicaciones claras de posibles comportamientos de evasión fiscal.Los resultados muestran que nuestro método propuesto puede mejorar en gran medida la eficiencia de la detección de evasión fiscal potencial, así como proporcionar explicaciones claras de posibles comportamientos de evasión fiscal.

## A Decision Tree and Naïve Bayes algorithm for income tax prediction

Resumen

Una de las preocupaciones con respecto al sistema de recaudación de impuestos es la selección incorrecta de casos. La selección manual de los casos de auditoría por parte de los auditores (cuya función es detectar casos individuales de incumplimiento tributario) en base a su conocimiento experto del comportamiento de los contribuyentes, no puede descubrir todos los patrones de comportamiento de incumplimiento ocultos en los datos históricos. Además, la selección aleatoria de casos de auditoría no se centra en los riesgos más altos. En otras palabras, la selección manual tiene un alto costo de oportunidad si se utiliza como único método de selección. La inteligencia computacional proporciona métodos, técnicas y herramientas, que se han enseñado a hacer predicciones precisas de impuestos sobre la renta basadas en observaciones anteriores. Los datos fueron recuperados de la situación ambiental en tiempo real. La aplicación de los métodos de inteligencia computacional demostró ser eficiente en el aprendizaje de un algoritmo de clasificación para clasificar a los contribuyentes que cumplen y no cumplen. El nuevo algoritmo fue evaluado y validado en pruebas empíricas en el mismo conjunto de datos. Aunque este algoritmo tenía la misma medida de rendimiento que el empaquetado, superó a los otros clasificadores múltiples existentes en términos de rendimiento. Esto ilustra un sistema automatizado que replica la operación de investigación de los auditores de riesgo tributario humanos.

Introducción

El desarrollo de modelos es una disciplina compleja. Sin embargo, desarrollar un modelo no se trata de fórmulas matemáticas complejas o diagramas de diseño complejos. Más bien, el proceso, en este caso relacionado con asuntos fiscales, se refiere al descubrimiento de patrones de **comportamiento** en los seres humanos y luego a la utilización de esos patrones para producir conocimiento en colaboración, que se puede traducir en un código ejecutable. En el reino animal, el **comportamiento.** Los patrones se pueden predecir razonablemente, pero esto no se aplica a los seres humanos. El concepto de ética debe ser observado y respetado. Algunos datos son altamente confidenciales y no son fácilmente accesibles u obtenidos. Algunos casos pueden producir falsos negativos (siendo de bajo riesgo mientras que en realidad son de alto riesgo) y no pueden abordarse hasta que haya evidencia concreta de que se debe realizar una auditoría, lo que a veces puede que nunca ocurra.

La selección aleatoria de contribuyentes que no cumplen con el conocimiento experto de los auditores sobre el **comportamiento** de los contribuyentes no se centra en el alto riesgo. En otras palabras, este método tiene un alto costo de oportunidad si se utiliza como único método de selección. El costo de oportunidad es el costo pagado al renunciar a una opción por otra: el costo de dejar pasar la siguiente mejor opción al tomar una decisión. La selección de casos basada en el riesgo identifica a los contribuyentes que tienen más probabilidades de no cumplir (Vellutini 2011**).**

El impuesto sobre la renta es una fuente importante de ingresos para los gobiernos de los países en desarrollo y los desarrollados (Hudson y Teera, 2004). La cantidad de ingresos generados por un gobierno a partir de impuestos para su programa de gastos depende, entre otras cosas, de la disposición de los contribuyentes a cumplir con las leyes fiscales de un país (Eshag 2006). El incumplimiento de las disposiciones fiscales sugiere que un contribuyente está cometiendo un acto de incumplimiento (Kirchler 2007). El incumplimiento tributario se produce por no presentar las declaraciones de impuestos, informar mal los ingresos o reportar las restas permisibles de los ingresos imponibles o impuestos debidos (Ayuba, Saad y Arifin 2015; Shome, Aggarwal y Singh 1996). La supervisión del cumplimiento fiscal implica la recopilación, el procesamiento y la interpretación de los datos relacionados con la condición de incumplimiento crítico **comportamiento**. El **comportamiento** compatible incluye el registro de impuestos, la declaración del ingreso apropiado, el pago de la deuda tributaria y la presentación de las declaraciones de impuestos.

La selección de casos basada en datos, en la que se basa la investigación de este documento, requiere criterios objetivos y no depende de la discreción del funcionario tributario. En su lugar, los métodos computacionales se utilizan para extraer información valiosa de los datos sobre el **comportamiento** del contribuyente. Los datos están capacitados para aprender una función de destino que se puede usar para predecir los valores de un atributo de clase discreta, por ejemplo, una declaración de impuestos correcta o una declaración de impuestos incorrecta y un riesgo alto o bajo.

Es necesario categorizar este riesgo en dos grupos: alto riesgo y bajo riesgo. Para poder clasificar este riesgo se requiere desarrollar un modelo. El objetivo de un esfuerzo de extracción de datos prescriptivo es automatizar un proceso de toma de decisiones mediante la creación de un modelo capaz de hacer una predicción mediante la asignación de una etiqueta. Una medida importante de dicho modelo sería su precisión (Berry y Linoff 2000). El objetivo de esta investigación, tal como se presenta en este documento, fue desarrollar dicho modelo. Se propuso un nuevo modelo de clasificador, el árbol de decisión y el algoritmo híbrido de Naïve Bayes (DTNBC), que utiliza alguna capacidad de optimización. El modelo en última instancia necesitaba ser capaz de hacer una predicción precisa.

El cumplimiento tributario aún está lejos de ser óptimo, por lo que la aplicación de las leyes tributarias aún es deficiente y tiene muchas lagunas. No existe un método científico que ayude a abordar estas lagunas de cumplimiento tributario. Este documento describe una propuesta para predecir las características de los contribuyentes de alto riesgo y cómo distinguirlos de las características de los contribuyentes de bajo riesgo. El modelo debe ayudar con la decisión futura de enviar un cliente para ser auditado. Para reiterar, el enfoque de este documento está en el proceso de determinar una función objetivo para poder predecir los valores de un atributo de clase discreta, por ejemplo, un contribuyente de alto riesgo o de bajo riesgo.

## A visual analytics system to support tax evasion discovery

Resumen

Este documento describe TaxNet, un sistema de apoyo a la decisión para el descubrimiento de la evasión fiscal, basado en un lenguaje visual poderoso y en técnicas avanzadas de visualización de redes. Se ha desarrollado en cooperación con la Agencia de Ingresos de Italia, donde se utiliza actualmente. TaxNet permite a los usuarios definir visualmente las clases de patrones sospechosos, explota tecnologías de coincidencia de patrones de gráficos efectivos para extraer rápidamente subgrafos que corresponden a uno o más patrones, proporciona facilidades para fusionar convenientemente los resultados e implementa nuevos índices de centralidad ad-hoc para clasificar Los contribuyentes en función de su riesgo fiscal. Además, ofrece una interfaz visual para analizar e interactuar con aquellas redes que coinciden con un patrón deseado. El documento analiza los resultados de un estudio experimental y algunos casos de uso realizados con expertos en datos reales y en un entorno de trabajo real. Los experimentos evidencian la efectividad de nuestro sistema.

Introducción

La evasión fiscal representa uno de los principales problemas de muchos gobiernos, debido a su fuerte impacto económico, político y social (véase, por ejemplo, Basta et al. 2009; Gonzalez & Velásquez 2013; Goumagias et al. 2012; Matos et al . 2015; Tian et al. 2016; Wu et al. 2012). Italia se encuentra entre los países particularmente afectados por este fenómeno: en el período 2007-2013, el gobierno italiano estimó un promedio anual de evasión fiscal de alrededor de 91.4 mil millones de euros. Como consecuencia, la Agencia de Ingresos de Italia (IRV a continuación) dedica cada año una gran cantidad de recursos humanos y económicos para contrastar el problema. Los funcionarios tributarios italianos tienen que lidiar continuamente con un número abrumador de fuentes de datos heterogéneas, gestionadas a través de diferentes aplicaciones de software.Esto causa redundancia de información y hace que sea extremadamente difícil para los analistas preservar su mapa mental durante el proceso de auditoría fiscal. Además, las aplicaciones de software actuales que se utilizan en el IRV se basan en un paradigma centrado en el contribuyente que no permite una fácil exploración de las relaciones entre diferentes sujetos. Por lo tanto, la evasión fiscal que involucra a grupos de sujetos, en lugar de individuos solos, suele ser más difícil de descubrir.

Este documento trata sobre el diseño de un nuevo sistema de apoyo a las decisiones para el descubrimiento de la evasión fiscal, capaz de ayudar a los analistas a superar los problemas anteriores. Nuestra contribución es doble:

**(yo)**                  Presentamos el sistema TaxNet, desarrollado en colaboración con el IRV. Apoya el trabajo de los funcionarios tributarios mediante un poderoso lenguaje visual y técnicas de visualización en red. Modela el conjunto de datos como una red unificada, cuyos nodos representan a los contribuyentes y cuyos bordes son diferentes tipos de relaciones económicas y sociales entre ellos. El usuario puede definir visualmente las clases de patrones sospechosos, basándose tanto en las propiedades topológicas como en los atributos de nodo / borde. T ax N et explota técnicas efectivas de comparación de patrones de gráficos para extraer de manera eficiente los subgrafos que corresponden a uno o más patrones sospechosos, proporciona facilidades para fusionar convenientemente los resultados e implementa nuevos índices de centralidad ad hoc para clasificar a los contribuyentes en función de su riesgo fiscal.El sistema también ofrece herramientas visuales para interactuar con aquellos subgrafos que coinciden con un patrón deseado, para obtener más detalles o para filtrar información menos relevante. Para ejecutar eficientemente las rutinas de coincidencia de patrones gráficos en redes grandes, los datos se almacenan convenientemente en una base de datos gráfica en lugar de una base de datos relacional tradicional.

**(ii)**                Evaluamos la validez de TaxNet en datos reales y en un entorno de trabajo real. A saber, el sistema se adopta actualmente en el IRV. Analizamos los resultados de un estudio experimental y dos casos de uso en datos manejados en el IRV para estimar la efectividad del sistema. Los datos experimentales sugieren que el uso de T ax N et tiene un fuerte impacto en el proceso de análisis de riesgo fiscal: reduce el tiempo necesario para ejecutar tareas de análisis fundamentales, facilita la recuperación de patrones sospechosos y aumenta la confiabilidad de los resultados.

El resto de este documento está organizado de la siguiente manera. La Sección 2 analiza el trabajo relacionado con nuestra investigación y destaca las diferencias y novedades de nuestro sistema con respecto a los enfoques anteriores. La Sección 3 ilustra ejemplos de patrones sospechosos o fraudulentos en el dominio de evasión fiscal y aclara la importancia de la visualización tanto para la definición de dichos patrones como para el descubrimiento de información adicional relacionada con los temas que los emparejan. La sección 4 recuerda las definiciones básicas sobre gráficos y redes, utilizadas en el documento. La Sección 5 resume el flujo de trabajo típico de una auditoría fiscal en el IRV, que nos guió en el diseño del sistema. En la misma sección, formalizamos el modelo de datos de TaxNet, según las consideraciones de la Sección 3. Las Secciones 6-7 describen el lenguaje visual de TaxNet y sus herramientas de análisis visual.Las Secciones 89 presentan los estudios experimentales y los casos de uso de TaxNet realizados en el IRV. La sección 10 concluye el documento y describe el trabajo futuro.

## Predicting tax avoidance by means of social network analytics

Resumen

Este estudio predice la evasión fiscal mediante el análisis de redes sociales. Extendemos la literatura anterior al ser los primeros en construir un modelo predictivo que incluye una mayor variación de las características de la red. Construimos una red de empresas conectadas a través de la membresía de la junta compartida Luego, aplicamos tres técnicas analíticas, regresión logística, árboles de decisión y bosques rand om; para crear cinco modelos utilizando las características de la empresa, las características de la red o las combinaciones diferentes de ambos. Un bosque aleatorio que incluye las características de la empresa, las características de la red de las empresas y las características de la red de los miembros de la junta proporciona el mejor rendimiento con un aumento mínimo de 7 pp en el AUC. Por lo tanto, la inclusión de efectos de red mejora significativamente la capacidad predictiva de los modelos de evasión de impuestos,lo que implica que los miembros de la junta exhiben un conocimiento específico que se puede transmitir a través de las empresas. Encontramos que tener miembros de la junta directiva sin conexiones con compañías de impuestos bajos reduce la probabilidad de ser una firma de impuestos bajos. De manera similar, cuanto más alta sea la tasa de impuestos promedio de las compañías a las que está conectado un miembro de la junta, menor será la posibilidad de ser de bajos impuestos. Por otro lado, estar conectado con más empresas de impuestos bajos aumenta la probabilidad de ser de bajos impuestos. De acuerdo con la literatura anterior sobre variables específicas de la empresa, el PP&E tiene una influencia positiva en la probabilidad de ser de bajos impuestos, mientras que el EBITDA tiene un efecto negativo. Nuestros resultados son informativos para las empresas en cuanto a la experiencia del director que desean atraer en sus juntas. Adicionalmente,Los analistas financieros y las agencias regulatorias pueden usar nuestros conocimientos para predecir qué empresas tienen probabilidades de tener impuestos bajos y estar potencialmente en riesgo.

Introducción

Existe una variación considerable en los impuestos que se pagan entre las organizaciones corporativas [20]. Si bien las empresas gozan de los beneficios de la evasión fiscal pagando menos impuestos, no está exento de riesgos, ya que las autoridades pueden imponer multas y sanciones por evasión fiscal, y la evasión fiscal puede implicar costos políticos y de reputación significativos [27]. Motivados por la variación en los impuestos que se pagan y las diferentes compensaciones por eludir impuestos, los investigadores deben examinar los factores determinantes de por qué las empresas se involucran en la evasión fiscal. En este contexto, muchos estudios se centran en las variables específicas de la empresa y los diversos incentivos que reciben los gerentes [2, 14, 33, 38]. Además, se pueden observar las diferentes variables de gobernabilidad, la calidad de los sistemas de información y los diversos tipos de experiencia en el consejo. o la oficina de auditoría, lo que indica que la planificación fiscal requiere un cierto nivel de experiencia [17, 27, 32]. Este documento utiliza técnicas del dominio de análisis de redes sociales para desarrollar un modelo de clasificación predictiva para la evasión fiscal. Motivados por los hallazgos recientes de que las empresas adquieren la experiencia requerida para la ejecución exitosa de estrategias corporativas a través de su red de directores [5, 19, 23, 25], nos enfocamos en varias características de red de firmas y directores y su validez predictiva para explicar la evasión fiscal . Hasta la fecha, no se sabe mucho sobre el papel de los vínculos sociales de los directores a través de las empresas y si comparten un conocimiento crucial que puede explicar la evasión fiscal (20, p. 146).Nos fijamos en cómo las firmas están conectadas a través de las juntas directivas compartidas y cómo el conocimiento compartido en la red a través de conexiones con firmas con impuestos bajos y firmas sin impuestos bajos puede ser informativo para hacer predicciones sobre si una firma podrá mantener (o convertirse) Una firma de bajos impuestos en el futuro. Mostramos que una combinación de las características de la empresa y las características de la red tanto de la empresa como de los miembros de su junta proporciona el mejor desempeño predictivo. Como tal, un modelo híbrido que incluye características tanto de la empresa como de la red (utilizando un bosque aleatorio) es capaz de identificar más empresas con impuestos bajos, y destaca la importancia de varias características de la red para predecir la evasión fiscal.Mostramos que una combinación de las características de la empresa y las características de la red tanto de la empresa como de los miembros de su junta proporciona el mejor desempeño predictivo. Como tal, un modelo híbrido que incluye características tanto de la empresa como de la red (utilizando un bosque aleatorio) es capaz de identificar más empresas con impuestos bajos, y destaca la importancia de varias características de la red para predecir la evasión fiscal.Mostramos que una combinación de las características de la empresa y las características de la red tanto de la empresa como de los miembros de su junta proporciona el mejor desempeño predictivo. Como tal, un modelo híbrido que incluye características tanto de la empresa como de la red (utilizando un bosque aleatorio) es capaz de identificar más empresas con impuestos bajos, y destaca la importancia de varias características de la red para predecir la evasión fiscal.

Nuestro estudio entrega varias contribuciones de investigación y gestión. En primer lugar, al incluir un extenso conjunto de funciones de red y al construir una red más compleja, podemos beneficiarnos de información más detallada sobre los miembros de la junta y sus relaciones con otras compañías. Esto mejora la comprensibilidad de nuestros modelos, ya que las características son bajas para una discusión del impacto de las características específicas de la red. Además, el rendimiento predictivo de nuestros modelos ha mejorado, lo que permite una mejor identificación de las empresas con impuestos bajos. En segundo lugar, contribuimos a la gestión y la sociedad. Nuestros resultados informan a la administración sobre la experiencia directa o la experiencia que desean atraer si desean una estrategia de impuestos bajos. Confirman que atraer directores que estén conectados a firmas de impuestos bajos ahora o en el pasado, puede afectar la tasa impositiva propia de las compañías,sugiriendo que estos rectores ofrecen un conocimiento crucial o tienen una experiencia valiosa para mantener una estrategia de impuestos bajos. Además, nuestros modelos predictivos pueden ayudar en la identificación a priori de las empresas que mantendrán una estrategia de impuestos bajos en el futuro y, como tal, brindarán información valiosa para los analistas financieros y las agencias reguladoras.

Nuestro artículo está estructurado de la siguiente manera. En primer lugar, en la Sección 2, analizamos la investigación relacionada con la evasión fiscal y el análisis de redes sociales para ilustrar la importancia y la novedad de nuestro estudio. A continuación, la Sección 3 describe nuestra metodología. Nuestros resultados se presentan en la Sección 4 y se discuten consecutivamente. Finalmente, la Sección 5 concluye nuestro estudio.

## Tax evasion in Sweden 2002–2013: interpreting changes in the rot/rut deduction system and predicting future trends Tax evasion in Sweden 2002–2013: interpreting changes in the rot/rut deduction system and predicting future trends

Resumen

En este artículo informamos sobre los efectos de los cambios recientes en la política fiscal en Suecia como un estudio de caso de evasión fiscal. Utilizamos los principios de prevención de la delincuencia situacional y los conocimientos extraídos de la teoría de la acción situacional para construir un marco analítico para evaluar cómo los cambios en la política fiscal afectan las motivaciones de los individuos y las empresas para la evasión fiscal; primero Disminuyendo las motivaciones y luego incrementándolas. El análisis se basa principalmente en datos secundarios de informes publicados por la Agencia Tributaria de Suecia. Luego hacemos predicciones específicas sobre las tendencias futuras en la evasión fiscal. Analizamos cómo los cambios en el sistema de deducción de impuestos pueden afectar a los compradores y proveedores de ciertos servicios que están gravados y predecimos que un porcentaje cada vez mayor de ellos considerará participar en la evasión fiscal. Concluimos discutiendo las acciones que las autoridades fiscales podrían tomar para prevenir aumentos en la evasión fiscal.

Introducción

El término "brecha fiscal" se refiere a la diferencia entre los impuestos que realmente se pagaron al gobierno en un período de reporte particular y lo que se debería haber pagado de acuerdo con las reglas determinadas por los controles de la Agencia Tributaria. En Suecia, varios indicadores sugieren que la brecha fiscal disminuyó a partir de finales de la década de 2000. La evidencia de esta reducción proviene de una serie de encuestas independientes sobre individuos y compañías, así como cálculos basados ​​en cuentas nacionales y el intercambio internacional de información (para una revisión, consulte Skatteverket [1]). Este cambio en el comportamiento (reducción de la evasión fiscal o mayor cumplimiento tributario) ha sido acompañado por cambios en la actitud de las personas hacia el cumplimiento tributario; por ejemplo, menos personas considerarían emplear mano de obra ilegal (The Swedish Tax Skatteverket [1]). La caída comenzó en 2007,Un año después se introdujeron las reformas fiscales. En efecto, estas reformas redujeron la cantidad de impuestos que los individuos y las empresas tenían que pagar cuando contrataban a trabajadores para determinados tipos de trabajo llamados en sueco, Rot / Rutavdrag. El trabajo de Rot cubre reparaciones, conversiones y extensiones, mientras que Rut incluye limpieza, mantenimiento y lavandería. En enero de 2016, se volvieron a cambiar las reglas fiscales, pero en la dirección opuesta, lo que hace que los impuestos sean más altos y el costo de la mano de obra sea más caro. Las reglas impositivas se cambiaron nuevamente, pero en la dirección opuesta, lo que hace que los impuestos sean más altos y el costo de la mano de obra sea más caro. Las reglas impositivas se cambiaron nuevamente, pero en la dirección opuesta, lo que hace que los impuestos sean más altos y el costo de la mano de obra sea más caro.

En este artículo, argumentamos que las reformas fiscales han tenido un impacto directo en las motivaciones de los individuos y las empresas con respecto al pago de impuestos y en sus actitudes generales hacia el cumplimiento de las leyes fiscales. Utilizamos los principios de prevención de la delincuencia situacional y los conocimientos extraídos de la teoría de la acción situacional para construir un marco teórico para evaluar cómo los cambios en la política fiscal en dos momentos (2006 y 2016) afectan las motivaciones de los individuos y las empresas para la evasión fiscal; Primero disminuyendo las motivaciones y luego incrementándolas. Comenzamos primero describiendo cómo el sistema de deducciones fiscales sobre los costos laborales afectó el atractivo del fraude fiscal en una dirección negativa (es decir, hizo más probable el cumplimiento tributario).luego, especulamos cómo el cambio en el sistema impositivo que entró en vigencia en enero de 2016 afectará tanto a los compradores como a los proveedores de estos servicios, por lo que estarán más dispuestos a considerar engañar al sistema. Luego predecimos que la evasión fiscal aumentará como resultado de los cambios en la política fiscal que hacen que la evasión fiscal sea más atractiva.

## Tax Fraud Detection for Under-Reporting Declarations Using an Unsupervised Machine Learning Approach

Resumen

El fraude fiscal es el acto intencional de mentir en un formulario de declaración de impuestos con la intención de reducir la obligación tributaria. La notificación incompleta es uno de los tipos más comunes de fraude fiscal, que consiste en completar un formulario de declaración de impuestos con una base impositiva menor. Como resultado de esta ley, los ingresos fiscales se reducen, lo que socava la inversión pública.

Introducción

El fraude fiscal es un fenómeno global que afecta a la sociedad en su conjunto. Este fenómeno se puede describir como un acto intencional de mentir en una declaración de impuestos con la intención de obtener un beneficio financiero ilegal y reducir la obligación tributaria [2, 26]. Estudios recientes [6, 8] han estimado que los gobiernos de todo el mundo pierden aproximadamente 500 mil millones de dólares anuales. Las pérdidas de ingresos fiscales son particularmente mayores en los países de ingresos bajos a medianos en las regiones de África subsahariana, América Latina y el Caribe y Asia meridional, que también son los países que dependen más de los ingresos fiscales para su planificación fiscal. Como resultado, estos países son los más afectados por la escasez de presupuesto, lo que limita el alcance de su inversión pública. Es crucial para los gobiernos adoptar estrategias rentables de detección de fraude fiscal para distinguir entre actividades fraudulentas y no fraudulentas, clasificando así a los contribuyentes o actividades fraudulentas y permitiendo a las autoridades tributarias tomar medidas para reducir el impacto del fraude [3, 5, 16, 28] ].

Las autoridades fiscales locales son responsables de diseñar políticas dirigidas a garantizar un horizonte de finanzas sostenibles. Un aspecto clave para lograr este objetivo es permitir la verificación adecuada y eficiente del cumplimiento de los contribuyentes. Además, son responsables de encontrar métodos y modelos efectivos y eficientes para seleccionar a los contribuyentes para las auditorías fiscales. Estas auditorías a menudo se desarrollan para determinar la consistencia entre el monto pagado por un contribuyente y el valor real a pagar.

Según Castellón en [5], las autoridades fiscales tradicionalmente han abordado el fraude fiscal a través de dos enfoques: la experiencia de los auditores y los sistemas basados ​​en normas. El primer enfoque consiste en seleccionar aleatoriamente declaraciones de impuestos y auditarlas según la experiencia, la intuición y el conocimiento de dominio de los auditores de impuestos. Este último utiliza metodologías basadas en sistemas basados ​​en reglas. Un sistema basado en reglas, como lo describen Baesens et al. en [4], a menudo se implementa en forma de un conjunto de reglas if-then que detectan casos de fraude. Estas reglas se desarrollan mediante un proceso engorroso, donde los auditores, después de una revisión exhaustiva, detectan un caso de fraude fiscal, generalizan sus características e incluyen una regla en la base de conocimiento de fraude fiscal. Una vez aplicado a los casos entrantes, activa una alerta o señal cuando se detecta un fraude. Sin embargo, Estas técnicas tradicionales tienen dos desventajas importantes: (1) Se basan principalmente en experiencias pasadas, por lo que no pueden descubrir nuevos mecanismos de fraude por sí mismas. (2) El juicio subjetivo de los expertos hace que las bases de conocimiento de los sistemas basados ​​en reglas sean costosas de construir, mantener y actualizar [4].

Un enfoque más reciente para detectar el fraude fiscal es a través del uso de técnicas de extracción de datos, que proporcionan mecanismos para extraer y generar conocimiento de volúmenes sustanciales de datos para apoyar la detección de comportamientos fraudulentos, mejorando el uso de recursos [3, 5, 9, 11, 28]. Además, esta es una de las aplicaciones mejor establecidas de minería de datos tanto en el gobierno como en la industria. Las autoridades fiscales en la mayoría de los países han adoptado con frecuencia técnicas de minería de datos para ayudarles a identificar a los contribuyentes que evaden las obligaciones [28]

Hay trabajos, en la literatura, que utilizan técnicas de aprendizaje supervisado debido al uso de datos etiquetados o datos asistidos por auditoría. También se han realizado esfuerzos sustanciales en el desarrollo de modelos para la detección de fraudes y la calificación de riesgos utilizando técnicas de extracción de datos. Sin embargo, la auditoría de las declaraciones de impuestos es un proceso lento y costoso, por lo que la consolidación de un conjunto histórico de datos etiquetados requiere un tiempo y recursos valiosos, por lo que el acceso a la información etiquetada es extremadamente limitado. Esto restringe la aplicabilidad de este tipo de modelos para la detección del fraude fiscal.

Estas cuestiones motivan el objetivo de este documento: una técnica de detección para la detección de declaraciones de impuestos que no reportan sin tener datos históricos etiquetados. La subinformación es un tipo de fraude fiscal caracterizado por llenar una declaración de impuestos con una base impositiva menor; o reclamando deducciones y exenciones que no son aplicables [19, 27].

La idea principal, aunque bastante evidente, nos permite construir un modelo flexible y eficaz: las declaraciones de impuestos similares deberían pagar una base impositiva similar, por lo tanto, este modelo se basa en grupos de declaraciones. Una vez que se agrupa la declaración de impuestos, se hace una estimación de la anomalía de la base impositiva declarada teniendo en cuenta el grupo de la declaración. Lo validamos utilizando el impuesto de Delineación Urbana como caso de estudio en colaboración con la Secretaría de Hacienda del Distrito ("Secretaría de Hacienda Distrital").

El resto del documento está estructurado de la siguiente manera. La Sección 2 presenta el trabajo relacionado en el contexto de fraude, riesgo de crédito y extracción de datos. La Sección 3 presenta una descripción de la técnica propuesta utilizada en nuestro enfoque para detectar pagos de impuestos fraudulentos sin la necesidad de datos marcados. La Sección 4 muestra estas técnicas aplicadas en un caso específico y los hallazgos más importantes. Finalmente, la sección 5 presenta las conclusiones de este trabajo.

## IRTED-TL: An Inter-Region Tax Evasion Detection Method Based on Transfer Learning

Resumen

La detección de la evasión fiscal desempeña un papel crucial en el tratamiento de la pérdida de ingresos fiscales. Se han realizado muchos esfuerzos para desarrollar modelos de detección de evasión fiscal mediante el aprovechamiento de las técnicas de aprendizaje automático, pero no han construido un modelo uniforme para diferentes regiones geográficas porque un amplio suministro de ejemplos de capacitación es un requisito previo fundamental para un modelo de detección eficaz. Cuando no se dispone de datos fiscales suficientes, el desarrollo de un modelo de detección representativo es más difícil debido a la distribución desigual de características en diferentes regiones. Los métodos existentes enfrentan un desafío al explicar y rastrear los resultados derivados. Para superar estos desafíos, proponemos un método de detección de evasión fiscal interregional basado en el aprendizaje por transferencia (IRTED-TL).que está optimizado para aumentar simultáneamente los datos de entrenamiento e inducir la interpretación en el modelo de detección. Explotamos el conocimiento relacionado con evasión en una región y aprovechamos las técnicas de aprendizaje por transferencia para reforzar las tareas de detección de evasión fiscal de otras regiones en las que faltan ejemplos de capacitación. Proporcionamos un marco unificado que aprovecha los datos auxiliares mediante un mecanismo de aprendizaje por transferencia y construye un clasificador interpretable para la detección de evasión de impuestos entre regiones. Las pruebas experimentales basadas en datos fiscales del mundo real demuestran que el IRTED-TL puede detectar evasores de impuestos con mayor precisión y mejor interpretabilidad que los métodos existentes.Explotamos el conocimiento relacionado con evasión en una región y aprovechamos las técnicas de aprendizaje por transferencia para reforzar las tareas de detección de evasión fiscal de otras regiones en las que faltan ejemplos de capacitación. Proporcionamos un marco unificado que aprovecha los datos auxiliares mediante un mecanismo de aprendizaje por transferencia y construye un clasificador interpretable para la detección de evasión de impuestos entre regiones. Las pruebas experimentales basadas en datos fiscales del mundo real demuestran que el IRTED-TL puede detectar evasores de impuestos con mayor precisión y mejor interpretabilidad que los métodos existentes.Explotamos el conocimiento relacionado con evasión en una región y aprovechamos las técnicas de aprendizaje por transferencia para reforzar las tareas de detección de evasión fiscal de otras regiones en las que faltan ejemplos de capacitación. Proporcionamos un marco unificado que aprovecha los datos auxiliares mediante un mecanismo de aprendizaje por transferencia y construye un clasificador interpretable para la detección de evasión de impuestos entre regiones. Las pruebas experimentales basadas en datos fiscales del mundo real demuestran que el IRTED-TL puede detectar evasores de impuestos con mayor precisión y mejor interpretabilidad que los métodos existentes.Las pruebas experimentales basadas en datos fiscales del mundo real demuestran que el IRTED-TL puede detectar evasores de impuestos con mayor precisión y mejor interpretabilidad que los métodos existentes.Las pruebas experimentales basadas en datos fiscales del mundo real demuestran que el IRTED-TL puede detectar evasores de impuestos con mayor precisión y mejor interpretabilidad que los métodos existentes.

Introducción

La evasión fiscal causa una gran pérdida de ingresos en China. El gobierno chino informó que la tasa de pérdida de ingresos fiscales en China era más del 22 por ciento [1]. Especialmente en los últimos años, las medidas de evasión fiscal se han vuelto más diversas y encubiertas en China. Muchas empresas utilizan instalaciones avanzadas, métodos contables y factores humanos para evadir las inspecciones de impuestos, lo que dificulta el trabajo de auditoría. Mientras tanto, con el rápido desarrollo de la economía, los datos fiscales han estado creciendo rápidamente. El número de registros comerciales relacionados con impuestos anuales es de hasta 1.000 millones, y el pico diario de estos registros es de hasta diez millones. Esa gran cantidad de datos genera una presión tremenda en el trabajo de auditoría fiscal.

Los gobiernos nacionales han tomado una serie de medidas para detectar la evasión fiscal. Las autoridades fiscales han adoptado tres medios para la detección de la evasión fiscal en su funcionamiento diario: selección manual de casos, selección de casos por computadora y selección basada en denuncias [1]. La selección de casos basada en computadora se basa principalmente en técnicas de aprendizaje automático, que extraen características relacionadas con la evasión de datos históricos para capacitación y obtienen un modelo que se puede utilizar en la auditoría fiscal. Por lo tanto, se considera un método semiautomático y de ahorro de mano de obra aplicable en la era del big data .

Sin embargo, los métodos tradicionales basados ​​en el aprendizaje automático tienen un problema concreto en la práctica, la detección de evasión de impuestos entre regiones. Los métodos de detección de evasión de impuestos basados ​​en el aprendizaje automático asumen una amplia oferta de ejemplos de capacitación como un requisito previo fundamental para construir un modelo efectivo de detección de evasión de impuestos en una región geográfica. Sin embargo, la anotación de los conjuntos de datos en el dominio de tributación tiende a ser costosa y requiere mucho tiempo. Además, un modelo de detección de evasión fiscal entrenado para una región específica puede tener un alto error de generalización cuando se aplica a otras regiones debido a diferentes condiciones económicas y sociales. Las diferentes regiones tienen diferentes distribuciones de características. La mayoría de los modelos estadísticos deben reconstruirse desde cero utilizando los datos de entrenamiento recién recolectados.Este es un gran desafío al desarrollar un modelo universal de detección de evasión fiscal para diferentes regiones con diferentes condiciones económicas y sociales. Por lo tanto, la construcción de un modelo de detección de evasión para una región con la ayuda de datos auxiliares de otra región se ha convertido en un tema importante y desafiante.

El aprendizaje por transferencia [2] es un método para utilizar el conocimiento adquirido al resolver un problema para resolver un problema diferente pero relacionado. Disminuye la necesidad de experiencia de expertos y reduce en gran medida la cantidad de datos etiquetados necesarios en un dominio de investigación objetivo. El aprendizaje por transferencia se ha aplicado ampliamente en la clasificación de documentos, reconocimiento de imágenes, reconocimiento de voz, descubrimiento de conocimientos y otros campos. Aplicando el aprendizaje por transferencia, podemos absorber el conocimiento auxiliar de una región de origen que posee datos de capacitación adecuados y aplicarlos a una región objetivo de análisis de etiquetas para aumentar el aprendizaje en presencia de diferencias regionales causadas por disparidades económicas y sociales. Se puede aplicar para la detección de evasión fiscal interregional.

Sin embargo, existen varios desafíos cuando se aplica la transferencia de aprendizaje a la detección de evasión de impuestos interregional.

Primero, ningún estudio existente sobre detección de evasión fiscal se basa en el aprendizaje por transferencia. Pocos trabajos han explorado cómo la transferencia de aprendizaje puede usarse en la detección de evasión de impuestos interregional. Los estudios actuales sobre el aprendizaje por transferencia son difíciles de aplicar directamente a la detección de evasión fiscal debido a los requisitos de alta precisión e interpretabilidad en el campo de la tributación.

Segundo, debido a las diferencias regionales, hay pocas características comunes entre las regiones, lo que causa dificultades en el proceso de transferencia. Incluso con características comunes, sus distribuciones de probabilidad marginal pueden ser bastante diferentes. Por ejemplo, la edad promedio de los representantes legales en las ciudades costeras de China es de 36 años, pero es 47 en las ciudades del interior de China.

Tercero, como lo presentaron Tian et al. [1], los resultados de los métodos basados ​​en el aprendizaje automático no se pueden explicar y son contraintuitivos. Casi todos los modelos de aprendizaje automático y los métodos de aprendizaje por transferencia son modelos de caja negra debido a la operación de mapeo de características, que son vulnerables a ataques de seguridad [3] [4] [5]. Hacer que el modelo sea interpretable es un tema importante para desarrollar un sistema de detección de evasión de impuestos sólido y estable.

En este documento, proponemos un método de detección de evasión fiscal interregional basado en el aprendizaje por transferencia (IRTED-TL) para superar los desafíos anteriores y realizar la detección de evasión fiscal interregional con alta precisión e interpretabilidad de sonido. Integra Transfer Adaboost (TrAdaBoost) [6], Transfer Component Analysis (TCA) [7] y LightGBM [8]. Especialmente, TrAdaBoost es un método de aprendizaje por transferencia basado en instancias, que garantiza la capacidad de transferencia del IRTED-TL. LightGBM es compatible con la interpretabilidad y la precisión del IRTED-TL. El TCA, un método de aprendizaje por transferencia basado en características, reduce la diferencia entre las regiones, lo que optimiza aún más el rendimiento del modelo. Por lo tanto, el IRTED-TL puede proporcionar un modelo eficaz y explicable para una tarea de detección de evasión de impuestos en una región objetivo.

En el IRTED-TL, extraemos características basadas en bosque aleatorio y divergencia de Kullback-Leibler (KL). El bosque aleatorio se utiliza para la extracción de importancia de la característica y la divergencia KL mide la similitud entre las distribuciones de la característica. Basándonos en estas características extraídas, asignamos las características con divergencia KL que excede un valor de umbral utilizando TCA. Luego, se adopta LightGBM para identificar si un contribuyente ha exhibido un comportamiento de evasión de impuestos basado en características asignadas. Revisamos de forma circular los pesos de la muestra en las regiones de origen y destino según el resultado de la clasificación aplicando TrAdaBoost. Para evaluar la efectividad del IRTED-TL, se realizaron experimentos basados ​​en los datos fiscales del mundo real de cinco regiones en dos provincias en China.Los resultados muestran que el IRTED-TL puede detectar evasores de impuestos con mayor precisión y mejor interpretabilidad que los métodos existentes.

El IRTED-TL es original y difiere sustancialmente de los métodos anteriores de detección de evasión de impuestos. Estos métodos normalmente utilizan técnicas tradicionales de aprendizaje automático y suponen una amplia oferta de ejemplos de capacitación como requisito previo para construir un modelo de detección de evasión fiscal en una región específica. El método propuesto en este documento adopta técnicas de transferencia de aprendizaje mediante el uso de datos auxiliares interregionales para aumentar el aprendizaje cuando faltan ejemplos de capacitación en una región objetivo. El IRTED-TL construye un clasificador interpretable para inducir la interpretación en el modelo de detección para que los resultados derivados puedan rastrearse. Las principales contribuciones de este documento se resumen a continuación:

* Proponemos un método novedoso para la detección de evasión de impuestos interregional considerando la ausencia de datos de capacitación en una región objetivo y la interpretabilidad de los resultados derivados.
* Proporcionamos un marco unificado que aprovecha los datos auxiliares aplicando el aprendizaje por transferencia y construye un clasificador interpretable para manejar los problemas de detección de evasión fiscal cuando faltan las muestras de capacitación disponibles en una región objetivo.
* Justificamos el rendimiento del IRTED-TL mediante la comparación con el trabajo existente basado en un gran conjunto de datos del mundo real con trece escenarios de transferencia en cinco regiones en dos provincias de China. Los resultados muestran que el IRTEDTL puede mejorar en gran medida la precisión de la detección de evasión de impuestos interregional y proporcionar una mejor interpretabilidad que el trabajo existente.

El resto del artículo está organizado de la siguiente manera. La Sección 2 proporciona una breve revisión sobre el trabajo relacionado. En la Sección 3, formulamos el problema que intentamos abordar y proporcionamos notaciones clave que se utilizan en el documento. Proponemos el IRTED-TL para la detección de evasión de impuestos interregional en la Sección 4. Describimos los resultados experimentales y proporcionamos análisis y discusiones en la Sección 5. Finalmente, se presenta una conclusión en la última sección.

## Taxpayer compliance classification using C4.5, SVM, KNN, Naive Bayes and MLP

Resumen

Los ingresos fiscales tienen un papel muy importante para financiar las finanzas del Estado. Para obtener los ingresos fiscales óptimos, las autoridades tributarias deben realizar una supervisión fiscal a los contribuyentes de manera óptima. Al utilizar el sistema de autoevaluación de impuestos que es un cálculo de los contribuyentes, pagar y reportar sus propias obligaciones fiscales agregadas con los datos de otras partes creará una gran cantidad de datos. Por lo tanto, las autoridades fiscales deben conocer de inmediato el incumplimiento del contribuyente para una auditoría adicional. Esta investigación utiliza el algoritmo de clasificación C4.5, SVM (Máquina de vectores de soporte), KNN (Vecino más cercano K), Naive Bayes y MLP (Perceptrón multicapa) para clasificar el nivel de cumplimiento de los contribuyentes con cuatro objetivos que los contribuyentes corporativos cumplen formalmente y materialmente requerido, los contribuyentes corporativos cumplen formalmente lo requerido,Los contribuyentes corporativos cumplen con los requisitos materiales y los contribuyentes corporativos no cumplen con los requisitos formales y materiales. Los resultados de la clasificación de cada algoritmo se comparan y el mejor algoritmo se elige según el criterio F-Score, Precisión y Tiempo empleado para construir el modelo mediante el uso del método TOPSIS difuso. El resultado final muestra que el algoritmo C4.5 es el mejor algoritmo para clasificar el nivel de cumplimiento de los contribuyentes en comparación con otros algoritmos.

Introducción

El impuesto es una fuente de ingresos de un país que tiene un papel muy importante en comparación con los ingresos de otras fuentes. Para las empresas, el impuesto a las ganancias como una deducción de los ingresos netos obtenidos por la empresa, de modo que existe una tendencia para que las empresas realicen la administración de los ingresos y la administración de impuestos [1]. El auto - sistema de impuestos de la construirá enorme de datos que es un reto para las autoridades fiscales de forma rápida para detectar contribuyentes que no cumplen para futuras investigaciones. Para saber el contribuyente que no cumple debe hacerse el método correcto, preciso y la selección de muchas variables que determinan el objetivo.

La comprensión del proceso comercial y la obligación de los contribuyentes ayudarán a elegir la variable correcta para determinar el éxito en esta investigación. Los tipos de obligación fiscal para los contribuyentes corporativos en Indonesia incluyen el Impuesto a la Renta, Artículo 21 / Artículo 26, Impuesto a la Renta, Artículo 23, Impuesto a la Renta, Artículo 26, Impuesto a la Renta, Artículo 25, Impuesto a la Renta, Artículo 29 e Impuesto sobre el Valor Añadido (IVA). Cada tipo de impuesto tiene un formulario de declaración de impuestos, la fecha de vencimiento y la fecha de vencimiento del pago. El formulario de declaración de impuestos se denomina Declaración de impuestos y se compone de Declaración de impuestos periódica y Declaración de impuestos anual. La declaración periódica de impuestos es una obligación fiscal que se realiza todos los meses y la declaración anual de impuestos es una obligación fiscal que se realiza todos los años.En la consolidación de la ley en la República de Indonesia, el número 6 de 1983 relativo a las disposiciones generales y los procedimientos tributarios, modificada por última vez por la Ley número 28 de 2007, mencionó que el contribuyente puede reparar la declaración de impuestos ha sido reportada mientras no ha sido auditado Hace que los datos de la administración tributaria siempre crezcan y se vuelvan muy grandes.

La clasificación de algoritmos C4.5, SVM, KNN, Naive Bayes y MLP se utilizan en esta investigación para determinar el nivel de cumplimiento de los contribuyentes corporativos que se distinguen en función de cuatro objetivos que deben cumplir formalmente los contribuyentes corporativos y que se requieren materialmente. Los contribuyentes cumplen con los requisitos materiales y los contribuyentes corporativos no cumplen con los requisitos formales y materiales. El algoritmo C4.5 es el mejor algoritmo para clasificar el nivel de cumplimiento del contribuyente porque tiene el valor de preferencia más alto en función del criterio F, la precisión y el tiempo necesario para construir el modelo en comparación con otros algoritmos.

## 

## An effective contrast sequential pattern mining ap-proach to taxpayer behavior analysis

Resumen

La minería de datos para el análisis del comportamiento de los clientes se ha vuelto cada vez más importante en los negocios, sin embargo, un análisis más profundo de las transacciones y los comportamientos secuenciales sería aún más valioso, especialmente en la industria de servicios financieros, como banca y seguros, gobierno, etc. En una aplicación comercial de recaudación de deudas tributarias en el mundo real, a fin de comprender la relación interna entre los comportamientos secuenciales de los contribuyentes (pago, presentación y acciones) y el cumplimiento de sus deudas, debemos encontrar los patrones de comportamiento secuencial de contraste entre cumplidores y no conformes. contribuyentes cumplidores. Los patrones de contraste (CP) se definen como el elemento del conjunto que muestran la diferencia / discriminación entre dos clases / conjuntos de datos (Dong y Li, 1999). Sin embargo, los métodos de extracción CP existentes que sólo se puede extraer patrones del conjunto de elementos, no son adecuados para la minería de patrones secuenciales, como el tiempo - ordenó transacciones en los comportamientos de los contribuyentes secuenciales. Hasta el momento, se ha realizado poco trabajo en la minería de patrones secuenciales de contraste (CSP). Por lo tanto, para abordar este problema, desarrollamos un enfoque de minería CSP, eCSP, mediante el uso de una estructura de árbol CSP efectiva, que mejora el Prefijo Span Tree (Pei et al., 2001) para patrones de contraste de minería. Proponemos algunos criterios de filtrado de heurística e interés, y los integramos en el árbol CSP a la perfección para reducir el espacio de búsqueda y encontrar patrones interesantes para el negocio también. El rendimiento del enfoque propuesto se evalúa en tres conjuntos de datos del mundo real. Además, utilizamos un estudio de caso para mostrar cómo implementar el enfoque para analizar el comportamiento de los contribuyentes. Los resultados muestran un rendimiento muy prometedor y un valor empresarial convincente.

Introducción

El análisis del comportamiento se centra en actividades humanas, como transacciones e interacciones, en entornos y entornos específicos. Se ha destacado cada vez más en muchos campos, como la gestión de las relaciones con los clientes [20], la búsqueda y el uso en la web [1, 3], la industria de servicios financieros, el gobierno [27], etc. El comportamiento del cliente desempeña un papel muy importante como fuerza impulsora interna para muchos problemas comerciales. Por lo tanto, es de vital importancia comprender profundamente el comportamiento del cliente para proporcionar servicios personalizados a grupos específicos de clientes [8]. Normalmente, el análisis tradicional del comportamiento del cliente se basa en atributos estáticos del cliente, es decir, datos demográficos. Sin embargo, el comportamiento secuencial, que no ha recibido mucha atención de investigación hasta la fecha, es mucho más valioso. Por ejemplo, en los negocios del mundo real, es esencial comprender la relación entre los comportamientos de presentación / pago (secuenciales) de los contribuyentes y su cumplimiento de sus deudas, lo que implica encontrar patrones de comportamiento secuencial de contraste entre los contribuyentes que cumplen y los que no cumplen.

Los patrones de contraste (CP), que son los conjuntos de elementos que muestran la discriminación entre dos clases / conjuntos de datos, pueden resolver un problema similar en los datos que no son de secuencia, como las comparaciones de exámenes médicos, análisis de comportamiento de compra, etc. [10].

Por ejemplo, en el análisis de abandono de clientes, hay dos conjuntos de datos: un conjunto de datos contiene registros de clientes que han batido, y el otro contiene clientes que aún están activos. Supongamos que cada conjunto de datos tiene 25,000 registros. Después de la extracción de CP, encontramos un patrón <HasPhoneNumber? = No, NumOfTenureYears = 3-5, Changed Sales? = Yes> (esto significa que un cliente no tiene un número de teléfono en el sistema, han pasado de 3-5 años desde que él / ella había comprado el producto y el cliente ha cambiado de proveedor), que tiene 1,470 registros / soporte de clientes cancelados y ningún soporte de clientes activos. Esto significa que los clientes que están cubiertos por este patrón todos batidos. Desde una perspectiva empresarial, esta es una regla muy útil que puede distinguir a los clientes de clientes abandonados y activos.

Se ha demostrado que el CP es muy explicable y eficiente en una comparación de dos clases y tiene una alta tasa de precisión de predicción [10, 12]. El concepto fue propuesto por primera vez por Dong et al. [10] y más tarde fue estudiado por [5, 12, 13, 24], etc. Sin embargo, (1) los métodos de minería de CP existentes pueden resolver problemas de minería de patrones de conjunto de elementos pero no la minería de patrones secuenciales. Dado que los datos de secuencia son una forma especial de datos en aplicaciones reales, como en comparaciones de proteínas, minería de texto, análisis de clics en la web, etc., los métodos actuales de minería de CP no son adecuados para minar patrones secuenciales. La minería de secuencias satisfaría más espacio de búsqueda enorme y una complejidad mucho mayor en comparación con los datos relacionales tradicionales o la minería de conjuntos de elementos. (2) Métodos de minería de patrones secuenciales existentes, es decir, FreeSpan [15], PrefixSpan [22], SPADE [26], SPAM [4], etc.no son adecuados para la minería de patrones secuenciales de contraste (CSP), ya que reducen principalmente el espacio de búsqueda por la propiedad antimonotona Apriori [2], que no es aplicable a la CSP [10]. Para abordar estos dos problemas, proponemos un enfoque de minería de CSP.

Es un reto minar CSP por las siguientes razones:

* En primer lugar, cuando el umbral de soporte es bajo cuando se usan métodos de extracción de patrones frecuentes existentes, el espacio de búsqueda crece enormemente.
* En segundo lugar, la propiedad Apriori es popular y es bastante efectiva para la poda en la minería de patrones frecuentes. Sin embargo, dado que no es válido para CSP, será muy difícil encontrar un enfoque eficiente para reducir el espacio de búsqueda en la minería de CSP.
* En tercer lugar, los enfoques de minería de CP anteriores a menudo producen una gran cantidad de patrones, lo que dificulta la elección manual de los interesantes en aplicaciones del mundo real. Por lo general, el procesamiento posterior se aplicaría para filtrar los que no son interesantes.

Para abordar esta necesidad, proponemos un nuevo algoritmo de minería CSP, que se llama eCSP. Puede extraer patrones secuenciales discriminatorios entre dos conjuntos de datos de secuencia, que muestran dos clases / tipos de población diferentes, para encontrar los patrones secuenciales que ocurren con frecuencia en un conjunto de datos, pero con poca frecuencia en otro.

Las principales aportaciones de este trabajo son las siguientes.

(1) Un algoritmo de extracción de CSP efectivo se basa en un árbol de CSP para buscar en todos los espacios de secuencia posibles para encontrar candidatos a CSP, mientras mantiene su información de clase correspondiente en las bases de datos proyectadas;

(2) Varios métodos de poda, basados ​​en algunas medidas de interés, se proponen e integran en el algoritmo para reducir significativamente el espacio de búsqueda y generar los patrones más interesantes directamente para abordar los problemas comerciales de la vida real.

(3) El algoritmo propuesto se examina utilizando tres conjuntos de datos del mundo real, los resultados muestran la alta eficiencia del enfoque y su rendimiento de escalabilidad robusta; y

(4) Implementamos un estudio de caso sobre la optimización de la recaudación de deuda fiscal mediante el uso de datos de comportamiento de secuencia de contribuyentes (es decir, transacciones de pagos, pagos y acciones) para determinar qué patrones de comportamiento pueden dar como resultado casos de deuda conformes, que no requieren ninguna acción para tomarse, con el fin de ahorrar recursos y maximizar el beneficio del negocio.

Este artículo está organizado de la siguiente forma: La sección 2 revisa el trabajo relacionado. La Sección 3 describe el problema de la minería de CSP y el análisis del comportamiento del cliente. La sección 4 detalla el enfoque eCSP. La evaluación experimental y un estudio de caso sobre el análisis del comportamiento de los contribuyentes se presentan en la Sección 5. La sección 6 concluye el documento y sugiere direcciones para futuras investigaciones.

## Inspecting compliance to many rules: An agent-based model

Resumen

El alcance y la complejidad cada vez mayores de las regulaciones y otras reglas que gobiernan la sociedad humana enfatizan la importancia de la inspección del cumplimiento de esas reglas. Los enfoques de uso frecuente para la inspección del cumplimiento adolecen de inconvenientes tales como suposiciones demasiado idealistas y estrechez de aplicación. Específicamente, los modelos de inspección con frecuencia se limitan a situaciones en las que la entidad inspeccionada tiene que cumplir con una sola regla. Además, las estrategias de inspección regularmente pasan por alto cierta información útil y disponible, como los costos variables de cumplimiento de las diferentes reglas.

Introducción

En 1788, en el federalista James Madison 2, se decía: "Si los hombres fueran ángeles, ningún gobierno sería necesario". Si los ángeles gobernaran a los hombres, no serían necesarios controles externos ni internos sobre el gobierno. Sin embargo, los hombres, como sabemos, no son ángeles.

La sociedad humana de hoy depende de una masa cada vez mayor de regulaciones, reglas y otras normas sociales. Los cambios en las regulaciones reflejan cambios en la sociedad y en nuestra comprensión de los procesos dentro de la sociedad, así como entre la sociedad y el medio ambiente. El aumento en la interconexión y la complejidad de la sociedad se refleja en el aumento en el alcance y la complejidad de las regulaciones, particularmente en áreas como la regulación financiera [1] y la regulación ambiental [2], y en particular en los EE . UU. Y la UE .

En las sociedades democráticas, las regulaciones típicamente introducen limitaciones y reglas de conducta que, al final, deberían ser beneficiosas para la sociedad en general. Sin embargo, para las personas y organizaciones que tienen que obedecer esas regulaciones, pueden parecer demasiado engorrosas, inútiles o incluso contraproducentes. Si surge una oportunidad, tanto las personas como las empresas podrían intentar violar las reglas o eludir sus responsabilidades, ya que, como señaló Madison, los hombres no son ángeles. Por lo tanto, por lo general, tiene que haber un perjuicio o una penalización para las entidades que no cumplen, para que las regulaciones sean efectivas. Debido a que las regulaciones generalmente no son autoimpuestas, requieren algún tipo de mecanismo externo de coerción, como las agencias de aplicación de la ley, agencias de inspección, etc.Las agencias de inspección y organizaciones similares por lo general desean lograr el máximo cumplimiento de las normas y regulaciones bajo su autoridad. Sin embargo, esas agencias no son omniscientes y, por lo general, no pueden saber si una entidad cumple o no, sin realizar algún tipo de procedimiento de inspección. Por otro lado, los procedimientos de inspección y los recursos disponibles rara vez son tales para permitir la cobertura total de todos los constituyentes. Por lo tanto, uno de los desafíos clave para cualquier organización de inspección es la selección óptima de entidades para la inspección. Este proceso de selección debe identificar, tan correctamente como sea posible, a los infractores y castigarlos. Además, el procedimiento de inspección debe servir como un elemento disuasorio para el comportamiento no deseado. Para complicar más las cosas,el creciente panorama regulatorio y el aumento de la complejidad y la cantidad de componentes se suele cumplir con la expansión correspondiente de los recursos de inspección. Por lo tanto, el aumento en la eficiencia del proceso de selección de inspección se convierte en primordial.

La selección de inspección y la inspección en sí han sido objeto de una extensa investigación científica. Hay varios enfoques al respecto, y cada uno tiene debilidades notables. En primer lugar, el análisis de los datos históricos ofrece información valiosa, pero no puede establecer causas y aislar variables. La experimentación en el mundo real es a menudo legalmente imposible o éticamente inaceptable. Se pueden realizar configuraciones de laboratorio y encuestas, pero a menudo surgen dificultades al intentar recrear configuraciones e incentivos del mundo real. Finalmente, a menudo se utilizan varios enfoques de modelación, pero también están plagados de deficiencias tales como suposiciones demasiado simplistas, perspectivas estrechas, aplicación limitada, capacidad de análisis analítica, falta de validación empírica, etc.

Este artículo describe varios enfoques de modelado para el análisis del problema de inspección y destaca sus ventajas y limitaciones. Además, el artículo presenta ICARUS (acrónimo: Inspección del cumplimiento de mis normas), un modelo basado en agentes para la inspección del cumplimiento de muchas reglas. El modelo describe un entorno genérico en el que una agencia de inspección inspecciona el cumplimiento de un conjunto de entidades a un grupo de reglas. Finalmente, el artículo demuestra un entorno de simulación simplificado del modelo y prueba la hipótesis de trabajo de que la realización de inspecciones basadas en el conocimiento de las necesidades de recursos para el cumplimiento reduce el incumplimiento total en el sistema.