

## MedicareVis: 面向医保反欺诈的联合可视分析方法

周杰辉<sup>1,2)</sup>, 朱融晨<sup>1)</sup>, 张玮<sup>1)</sup>, 陆俊华<sup>1)</sup>, 应豪超<sup>3)</sup>, 吴健<sup>3)</sup>, 陈为<sup>1)\*</sup>

<sup>1)</sup> (浙江大学 CAD&CG 国家重点实验室 杭州 310058)

<sup>2)</sup> (之江实验室 杭州 311121)

<sup>3)</sup> (浙江大学公共卫生学院 杭州 310058)

(chenvis@zju.edu.cn)

**摘要:** 医保欺诈造成严重经济损失, 对医保制度的安全稳定造成巨大冲击. 然而, 现有工作难以支持对多类欺诈行为的联合分析和探索. 基于多维时序相关的医保数据, 提出面向医保反欺诈的联合可视分析方法. 该方法支持医保数据的时空多角度过滤, 以更快地定位欺诈; 并通过不同类别、不同主体的欺诈之间的关联分析, 挖掘出隐蔽的欺诈行为; 与医保领域专家合作, 设计并实现一个面向医保反欺诈的可视分析系统 MedicareVis, 通过真实医保数据上的实例研究与专家访谈, 证明该方法在帮助检测欺诈行为关联上的有效性和实用性.

**关键词:** 可视分析; 医疗保险; 欺诈检测

中图法分类号: TP391.41      DOI: 10.3724/SP.J.1089.2021.18981

## MedicareVis: a Joint Visual Analytics Approach for Anti-Fraud in Medical Insurance

Zhou Jiehui<sup>1,2)</sup>, Zhu Rongchen<sup>1)</sup>, Zhang Wei<sup>1)</sup>, Lu Junhua<sup>1)</sup>, Ying Haochao<sup>3)</sup>, Wu Jian<sup>3)</sup>, and Chen Wei<sup>1)\*</sup>

<sup>1)</sup> (State Key Laboratory of CAD&CG, Zhejiang University, Hangzhou 310058)

<sup>2)</sup> (ZhejiangLab, Hangzhou 311121)

<sup>3)</sup> (School of Public Health, Zhejiang University, Hangzhou 310058)

**Abstract:** Medical insurance fraud causes serious economic losses, which has a great impact on the safety and stability of the medical insurance system. However, existing work does not support the joint analysis and exploration of various types of fraud. Based on the medical insurance data related to multi-dimensional time series, a visual analytics approach for anti-fraud in medical insurance is proposed. It can perform spatio-temporal filtering of medical insurance data, locate fraud quickly, and discover hidden frauds by performing the correlation analysis between different types and different subjects of fraud. We design and develop MedicareVis, a visual analysis system for medical insurance anti-fraud. We demonstrate the usefulness and effectiveness of our approach in helping detect the association of fraud through a case study on real-world medical insurance data and interviews with domain experts.

**Key words:** visual analytics; medical insurance; fraud detection

收稿日期: 2021-03-10; 修回日期: 2021-03-15. 基金项目: 国家重点研发计划(2019YFB1404802). 周杰辉(1997—), 男, 博士研究生, 主要研究方向为可视化; 朱融晨(1998—), 女, 硕士研究生, 主要研究方向为可视化; 张玮(1988—), 女, 硕士研究生, 主要研究方向为可视化; 陆俊华(1993—), 男, 博士研究生, 主要研究方向为可视化; 应豪超(1991—), 男, 博士, 助理研究员, CCF 会员, 主要研究方向为医疗健康大数据挖掘与分析; 吴健(1975—), 男, 博士, 教授, 博士生导师, CCF 会员, 主要研究方向为医学人工智能; 陈为(1976—), 男, 博士, 教授, 博士生导师, CCF 会员, 论文通讯作者, 主要研究方向为可视化、可视分析、大数据分析和人机混合智能.

医保数据是患者去医疗机构就诊开药和报销结算期间产生的数据集合<sup>[1]</sup>, 其具备大数据典型的 5V(volume, variety, velocity, veracity, value)特点<sup>[2]</sup>, 即数据体量大、数据种类多、时效性要求高、真实性强、价值密度低。通过对医保数据的分析, 可以帮助国家相关机构进行药品定价、政策制定等。

随着医疗保险覆盖面和基金规模的不断扩大, 以及定点机构的逐渐增加, 医保诈骗问题也日趋严峻。据国际透明组织统计, 全球卫生市场每年因欺诈损失数十亿美元<sup>[3]</sup>。因此, 有效挖掘欺诈行为对于减少经济损失、健全医保制度具有重要意义。

然而, 现有的挖掘方法常使用单一欺诈检测模型。如何高效地分析欺诈行为的关联, 发现欺诈者的潜在联系, 进而辅助审计和决策, 是亟须解决的问题。目前还面临许多挑战: (1) 实体关系属性复杂。医保牵扯多方实体, 如参保人、医生、医院和药品等, 且数据量大属性维度高, 这些都增加了分析难度。(2) 模型关联性不足。单一的欺诈检测模型只针对特定种类的欺诈行为, 难以挖掘涉及多类欺诈的复杂模式。(3) 可视交互功能缺乏。医保审计人员只能通过现有系统提供的简单统计功能对数据进行处理和人工分析, 这存在耗时长、检索效率慢的问题。尚不存在一个针对医保欺诈的可视分析系统, 支持专业人员实现敏捷交互、检索和分析。

本文总结了医保领域专家对于医保欺诈行为的分析和探索需求, 提出采用可视分析的范式, 将医保审计人员的领域知识与机器智能融合, 提升欺诈行为挖掘的效率。本文设计并实现一个交互式医保可视分析系统 MedicareVis, 融合多个特定类别的医保欺诈检测模型, 支持对医保时空数据的联合可视分析, 帮助专家快速探索和调查欺诈行为的关联关系。

## 1 相关工作

### 1.1 医保数据可视化

数据可视化技术是探索和理解海量多维复杂医保数据的重要工具<sup>[4]</sup>。倪明选等<sup>[5]</sup>指出, 具备良好可视化与互动功能的界面可以帮助专家有效地发掘医疗大数据潜在价值。Toyoda 等<sup>[6]</sup>从时间、保险公司、疾病和年龄 4 个角度设计可视化视图, 降低对原始大量医保索赔数据处理难度, 方便用户从全局和局部等不同侧重点对数据分析操作。Lin

等<sup>[7]</sup>利用医保记录中患者的就诊时间线可视化加快医生的诊断速度。Mohammed 等<sup>[8]</sup>构建了一个医疗领域安卓移动应用程序, 提供对就诊数据的可视分析功能。Bashyam 等<sup>[9]</sup>以临床异常问题为导向, 设计研发了一套“数字人”与临床信息可视化系统, 对诊疗过程中的数据进行全面展示。Ristevski 等<sup>[10]</sup>则指出通过元数据和数据流视图等可以帮助合规性的审计跟踪。彭燕妮等<sup>[11]</sup>对医保中常见时序数据的可视化方法进行了总结。纪守领等<sup>[12]</sup>则说明可视化方法有助于解释模型, 提高人工分析与决策的效率。

上述工作主要对基本的医保就诊和基金数据提供简单可视化图表和医保数据概览。本文针对欺诈行为的时空特性, 提供丰富的可视化设计与交互操作, 满足用户对欺诈行为深入检测和分析需求。

### 1.2 医保欺诈检测方法

随着医保数据的大量积累, 数据挖掘和神经网络模型是医保欺诈行为检测的主流技术。Viveros 等<sup>[13]</sup>最早提出将数据挖掘中的关联规则方法用于医保欺诈检测。Kumar 等<sup>[14]</sup>在保险公司处理医疗保险申诉行为时使用数据挖掘技术进行预测和预防支付错误, 从而降低医疗保险的不合理支出。2013 年, Skiba<sup>[15]</sup>通过对保险欺诈案的深入调查, 利用解释性的定性现象, 探索打击保险欺诈工作中的障碍。Ekina 等<sup>[16]</sup>将医学知识、不确定性量化规则引入贝叶斯方法, 从而识别欺诈的医生和病人。2016 年, Johnson 等<sup>[17]</sup>提出一种多阶段方法检测健康保险索赔中的欺诈, 并基于决策树模型计算风险阈值。Shin 等<sup>[18]</sup>提出一种评分模型自动化检测出现欺诈的诊所。虽然数据挖掘方法可以计算出欺诈行为, 但是还需要医保审计人员对数据进行额外的识别与确认。

Hubick<sup>[19]</sup>首次将简单感知机神经网络引入医疗欺诈识别。He 等<sup>[20]</sup>使用多层 BP(back propagation)神经网络模型提高医保欺诈检测的准确率。Liou 等<sup>[21]</sup>在中国台湾健康保险数据上对神经网络的欺诈检测效果进行实证研究。然而, 深度学习方法需要大量医保欺诈标注数据, 计算过程复杂, 难以对挖掘结果进行解释, 导致目前应用范围较为局限。

不同于其他工作, 本文专注于多类欺诈行为关联的任务, 针对欺诈检测模型和海量多维的医保数据设计可视编码方案, 以支持对欺诈行为的交互式挖掘与关联分析。

## 2 需求设计

通过与医保领域经验丰富的产品经理和医院医保研究人员共同探讨, 本文制定了针对欺诈行为的可视分析设计要求. 具体而言, 以访谈、周会与原型系统设计迭代方式, 结合数据实际属性特征, 深入剖析医保欺诈事件, 最终总结出3点需求和3种较为常见的欺诈行为.

### 2.1 需求

(1) 快速筛选欺诈行为. 对时间、空间和模型参数等进行快速调整, 便于用户筛选出感兴趣的欺诈行为.

(2) 关联多种欺诈行为. 考虑到欺诈行为的多样性和复杂性, 为了更深入地挖掘欺诈群体, 需要对多种类型的欺诈行为进行快速关联分析.

(3) 探索可疑群体的详细信息. 支持概览与细节的互动探索, 呈现欺诈关联证据, 为审计人员提供怀疑和风险评估的背景, 便于对数据进行深入检查与决策.

### 2.2 欺诈行为

(1) 群体就医. 参保人以群体方式多次在同一天同一家机构就诊的行为(排除慢病、规定病种、中药等具有固定就诊和开药周期的情况).

(2) 多地开药. 医生出借医师执照给其他不具有行医资格的人, 常表现为短时间内单个医生有多地开药记录.

(3) 虚假住院. 参保人故意利用住院方式报销一些非医保目录项目以及获得高报销比或以“医养”的方式长期多次违规住院的行为. 例如, 体检式住院、购药式住院和医养住院等, 常表现为药占比、检查比、住院费用等指标与正常值有较大偏离.

## 3 MedicareVis

### 3.1 系统框架

为满足医保欺诈行为的快速有效识别, 本文采用模块化开发方式研发 MedicareVis. 该系统后端采用 Django 开发, 用于数据预处理和数据分析; 前端包括可视分析模块, 采用 Vue 和 D3 开发, 支持多视图的交互与联动. 具体如图 1 所示.

数据预处理模块对原始数据进行清洗过滤. 由于数据表中存在无关冗余的字段, 且具有空值和不一致的情况, 通过数据预处理, 有助于为后续数据分析方法提供高质量的就诊时序数据, 提高欺诈检测效率.

数据分析模块允许用户对模型参数和分析时间段进行调整, 不同的用户可以根据审计目标查看某一类或多类模型的结果, 通过改变参数, 调整模型的效果, 增强模型的透明性和可理解性.

可视分析模块直观呈现模型检测的不同类别欺诈的就诊数据, 并通过医院分布地图视图和细节视图深入展示欺诈行为的就诊上下文, 审计人员可以从这些细节信息中提取出相应的欺诈证据.

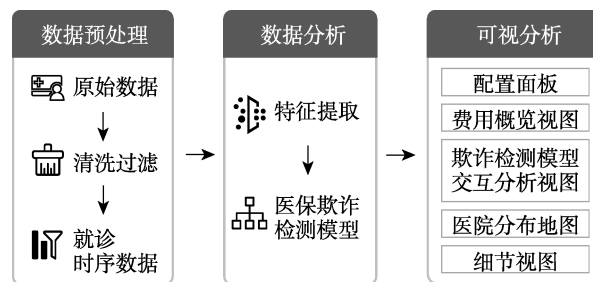


图1 本文系统框架

### 3.2 数据预处理

脱敏处理过的原始数据主要包括以下3种类型的数据库表, 其规模如表1所示.

表1 原始数据表信息

表格	记录数	字段数
参保人就诊资料	203 136	118
处方明细数据	1 718 227	91
费用结算	209 883	79

原始数据中主要涉及3类群体, 分别为参保人  $P$ 、医生  $D$ 、医院  $H$ , 每个实体均具备多种属性, 如参保人编号  $p$ 、年龄  $a$ 、医师编码  $d$ 、医院编号  $h$  等, 可用高维向量  $\{f_1, f_2, \dots, f_k\}$  进行表征.

原始数据经过缺失值过滤等处理后得到就诊时序序列  $S = \{s_1, s_2, \dots, s_m\}$ . 单次就诊  $s_i$  是一个高维向量, 包含多种属性, 如参保人的就诊日期、疾病类型、药品编号、就诊流水号和各种费用等信息.

### 3.3 数据分析

#### 3.3.1 群体就医

为了发掘经常同时同地就诊的群体就医欺诈行为, 需要找到参保人的频繁就诊模式, 在一段时间内符合该模式的参保人即有群体就医欺诈嫌疑.

形式化地, 在时间段  $[t_a, t_b]$  (单位: 天)内, 若人数为  $C$  的群体  $G$  中每个参保人的就诊序列  $S$  均包含模式  $M = \{(t_a, h_a), \dots, (t_b, h_b)\}$ , 则这些参保人为群体就医者. 人数  $C$  也称为该模式的支持度.

群体就医欺诈检测模型的输入为时间段和就

诊序列(过滤慢病、规定病种和中药),本文采用 Eclat 频繁模式挖掘算法<sup>[22]</sup>,与传统的 Aprior 算法<sup>[23]</sup>和 FP-Growth 算法<sup>[24]</sup>相比, Eclat 算法只需要扫描数据 2 次,并且不需要重复构造频繁模式(frequent pattern, FP)树,效率高,最终得到频繁模式和符合该模式的参保人就诊流水信息。

### 3.3.2 多地开药

医师执照如果被滥用或冒用,就会出现在短时间内,多家机构有同一名医师开药的记录。从参保人的就诊序列  $S$  中构建出医师的开药记录  $Q = \{q_1, q_2, \dots, q_R\}$ , 在时间段  $[t_a, t_b]$  内, 单个医师的开药记录  $q_s = \{(t_a, h_a), \dots, (t_b, h_b)\}$  中的多地开药违规记录为  $\{(t_i, h_i) | \exists t_j \in [t_a, t_b] \wedge h_i \neq h_j\}$ 。

由于有多机构坐诊的情况,算法限制多地开药时长  $t_b - t_a \in \{5, 10, 60\}$  (单位: min)。对于每位医生,利用滑动窗口遍历其开药序列,同时用多重集合维护窗口内的医疗机构编号,当窗口移动时,从多重集合中删除一个移出机构的编号,增加一个移入机构的编号,根据集合总大小和其中某个机构编号的数量是否相等,即可判断该时间阈值内是否存在多地开药行为。

多地开药检测模型输入为就诊时序序列和多地开药时间间隔阈值,经过医师开药记录的构建与违规行为挖掘,最终输出医师的违规开药记录和相应的参保人就诊流水信息。

### 3.3.3 虚假住院

为了确定与虚假住院欺诈类型相关的属性特征,本文与领域专家进行讨论,并对虚假住院实例进行分析,最后得出 7 个需要检测的指标,分别为住院天数、总住院次数、药占比、检查比率、住院总列支费用、住院日均次费用和年龄。

参保人可形式化表示为  $P = \{p_i, f_1, f_2, \dots, f_7\}$ ,  $f_i$  分别代表一段时间内参保人上述 7 个指标的数值,  $i = 1, 2, \dots, 7$ 。

由于缺乏有标记的异常数据,无法准确表示异常类别的潜在分布,本文采用了局部异常因子(local outlier factor, LOF)模型<sup>[25]</sup>,选择这个模型是因为它具有很多优点:(1)它是无监督学习模型,适用于本文的场景,即事先不知道欺诈者;(2)它为每个参保人分配了异常分数而不是二进制标签,有助于在可视分析中减少搜索空间;(3)该算法十分高效,其时间复杂度为  $O(M \ln N)$ ,其中,  $N$  为参保人的数量,因而该模型可以较为容易地集成到可视分析系统当中。LOF 定义为

$$F_k(P_i) = \frac{\sum_{y \in N_k(P_i)} D_k(y_i)}{|N_k(P_i)| D_k(P_i)} \quad (1)$$

其中,

$$D_k(P_i) = \frac{|N_k(P_i)|}{\sum_{q \in N_k(P_i)} (\max(d_k(q), d(P_i, q)))} \quad (2)$$

为  $P_i$  的邻域密度;  $N_k(P_i)$  为参保人  $P_i$  在特征空间上的  $k$  近邻;  $d(P_i, q)$  为  $P_i$  与其邻居  $q$  的欧几里得度量;  $d_k(q)$  为  $q$  与其  $k$  近邻的最远距离。直观地说,  $F_k(P_i)$  是基于参保人行为在特征空间中形成几个潜在聚类的假设而设计的,因此,  $P_i$  的离群因子可以通过仅将其与特征空间中的邻近特征向量而不是整个群体进行比较来确定。

虚假住院检测模型的输入为分析的时间段、就诊时序数据,以及计算得到的 7 个分析特征,考虑到实际数据中大多数为正常记录,所以默认取 LOF 异常得分前 1%~3%(用户可对该参数进行调整)的数据作为输出的虚假住院异常就诊记录。

## 3.4 可视分析

MedicareVis 系统通过 5 个协调的视图交互地挖掘和探索可疑的欺诈行为,如图 2 所示。

(1) 配置面板。由时间区间、机构类型和模型参数等组成。用户可以为算法配置相应的输入参数如群体就医最少人数、多地开药时间间隔阈值等。

(2) 费用概览视图。包括医保费用折线图和增长率日历热力图,可以引导审计人员找出他们关注的时间段,以便更快地检测到欺诈行为。

(3) 欺诈检测模型交互分析视图。平行坐标可视化不同类别的欺诈就诊行为,从左至右,坐标轴分别代表日期、参保人、医生、医院和欺诈类别,下方的散点图矩阵反映了它们两两之间的分布关系,散点的颜色编码欺诈类别,用户可以在平行坐标或散点图矩阵中选择其感兴趣的区域,其余图表会同步过滤和高亮显示。

(4) 医院分布地图。将用户感兴趣的欺诈交易数据按就诊机构的地理位置映射到地图上,地图支持缩放和拖拽,当地图代表的尺度增大时,相距较近的机构会被聚合在一起,以减少视觉混乱。异常总次数排名前 10 的机构使用红色进行突出显示,用户可以进一步单击机构,查看其涉及各个欺诈行为的具体指标,例如,群体就医处方相似度以及多地开药次数随时间的累计堆叠柱状图。

(5) 细节视图。图 2 中 E 区域所示为用户在图 2 中 C 区域和 D 区域中选择的相关医疗机构、医生

和参保人的具体细节. 饼状图展示列支费用和药品费用分布的各个细分项所占的比例, 柱状图从就诊

类别、人次、金额等多个维度对服务人群的结构信息进行查询, 以使用户进一步找出欺诈的证据.

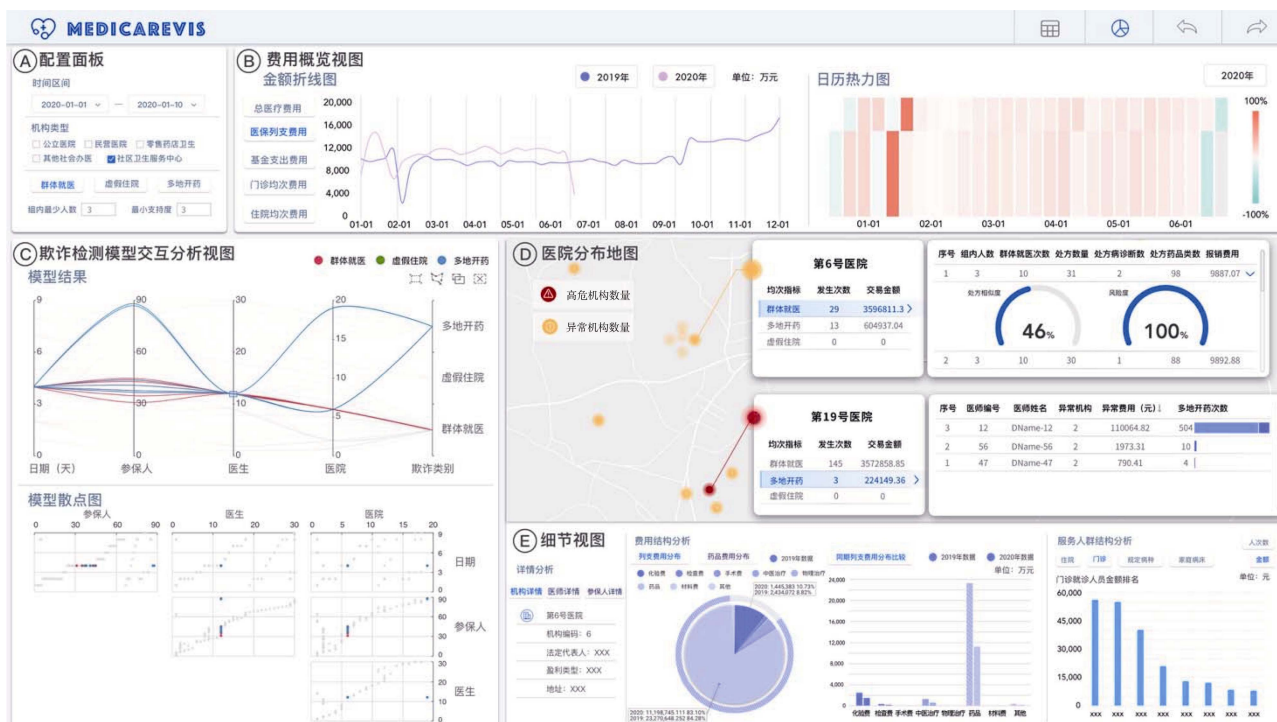


图 2 本文系统界面

### 4 实例

通过一个真实案例, 说明 MedicareVis 可以帮助发现和探索欺诈行为以及它们之间的关联关系.

一位专家对模型交互与可视分析的集成感兴趣, 想要找出欺诈行为之间的关联. 因此, 为了有一个总体的印象, 他首先在配置面板中选择了全部时间段(2019-01—2020-06), 机构类型为社区卫生服务中心; 接着在概览视图中单击“医保列支费用”标题卡, 发现 1 月上旬的支出费用有较大幅度增加, 这种突变立即引起了他的注意, 他把配置面板的时间段缩短为 2020-01-01—2020-01-10. 在运行 3 种欺诈检测模型后, 可疑的欺诈就诊记录出现在模型关联交互分析视图中, 如图 3 和图 4 所示.

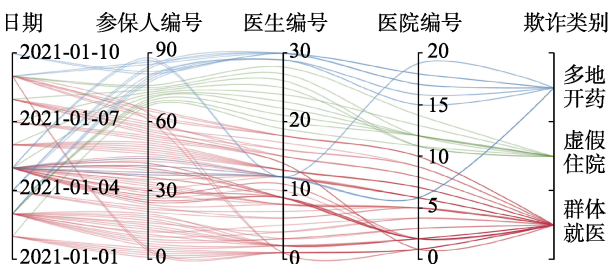


图 3 多类欺诈各主体的平行坐标

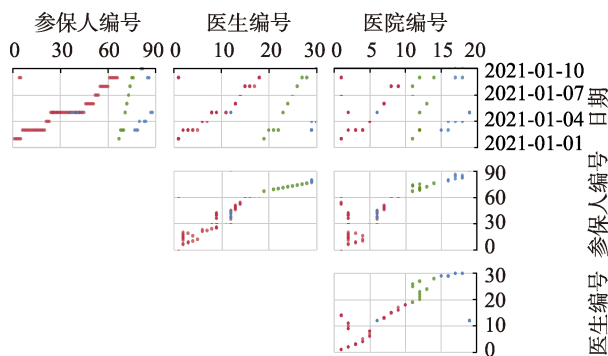


图 4 关联分析散点图矩阵

平行坐标轴中的一条线即为一次就诊记录, 坐标轴上线的疏密程度反映涉及的欺诈行为数目的多少. 专家注意到 12 号医生有较多的欺诈行为, 所以他选择该名医生, 高亮的视觉线索使专家追踪相关细节, 发现该名医生涉及多地开药和群体就医 2 种类型的欺诈, 专家进一步查看散点图矩阵, 在参保人-医生散点图中, 代表群体就医的红色点多呈现离散的柱状结构, 表明医生与患者有较强的相关性, 因此专家怀疑 12 号医生可能与对应的参保人串通进行骗保.

为了进一步确认, 专家在地图上选择了这个可疑团体就诊时所在的机构, 发现其欺诈涉及的 2

家机构(6 号和 19 号医院)的违规次数分别达到了 42 次和 148 次。结合地理位置,专家分析到这 2 家机构均属于同一片区的社区医院,医生之间都非常熟知,可为多地开药提供便利。专家进一步单击地图上代表机构的点,弹出的页面描述了不同类型欺诈的发生次数与金额等详细信息,如图 2 中 D 区域所示。专家通过排序发现在这 2 家医疗机构中,12 号医生的违规金额和次数均排名第 1,并且

与选中群体中的参保人有多次就诊往来。

专家接下来导航到细节视图,观察具体的就诊费用构成情况,如图 5 所示,该医生主要服务在门诊,费用构成主要包括西药费和中草药费。专家注意到 37 号参保人与 12 号医生的 2 种类别的欺诈行为均相关,在 6 号医院中,该参保人主要报销西药费,而在 19 号医院中,该参保人主要报销中草药费,这与 12 号医生的主要费用构成一致。

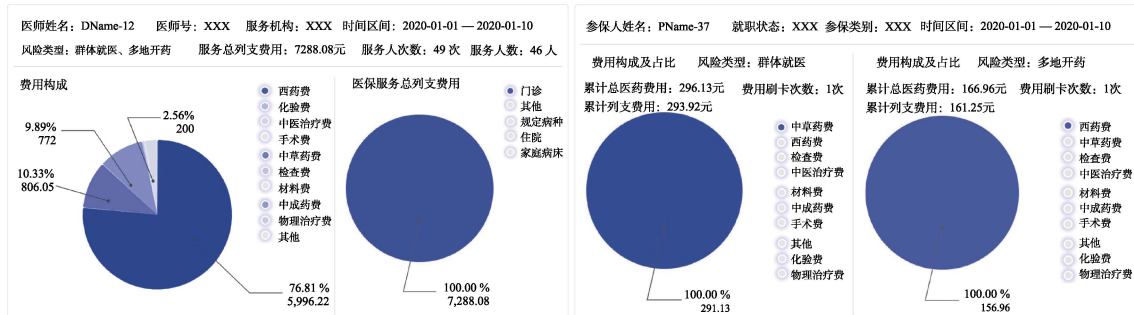


图 5 选中的医生和参保人的费用构成情况

这些短时间内具有较大药品差异的就诊行为引起了专家的怀疑,他们指出该名参保人很可能与该名医生串通勾结,通过收集多张医保卡在这里进行“套刷”以骗取医保基金,并通过在多家医院报销不同的药品掩盖欺诈行为。专家认为该欺诈团伙值得进一步调查,传统的审计方法很可能会忽略这一群体,本文系统为分析关联的欺诈行为并发现隐蔽的欺诈团伙提供了有效的方法。

## 5 专家评估

本文通过与 4 名专家的结构化访谈验证系统的有效性和实用性。各访谈环节持续 1 h,前 20 min 向专家介绍系统和交互形式;接下来 30 min 专家遵循放声思考法,亲自使用系统探索挖掘可疑的欺诈行为;最后 10 min 对系统进行评价,并对系统有效性和实用性进行问卷打分(1=非常高效/可用,5=非常低效/难用)。

(1) 欺诈检测方法。专家们均认为,以欺诈行为关联为导向的交互分析,有可能发现在目前审计过程中被忽略的欺诈团伙。第 2 位专家回顾已侦破的医保欺诈案件,根据经验,特定的欺诈团伙之间可能使用多种欺诈方法获取不正当的利益,通过在平行坐标中用视觉编码展示这种信息,能够允许快速导航和缩小搜索空间。另外,详细视图以就诊明细数据作为背景,能够帮助发现欺诈交易的证据。

(2) 系统和改进。所有专家对于系统有效性的平均评分为 2.1,实用性的平均评分为 1.3,这表明专家认为 MedicareVis 易于学习和高效分析。第 1 位专家认为系统对欺诈行为的关联识别十分有用且高效,能够让领域专家参与到决策审计的过程中。另外,第 4 位专家补充,医保支出费用的变化可能是宏观经济和季节性疾病等因素影响的结果,而不是由于欺诈行为,这需要更多的数据来验证。

## 6 结语

### 6.1 讨论和未来工作

实例分析和专家访谈证明了 MedicareVis 系统的有效性和实用性,但该系统还有一些方面值得进一步讨论说明。

(1) 可扩展性与模型精度。由于屏幕尺寸有限,MedicareVis 在平行坐标中展示大量的欺诈行为时会导致视觉混乱。用户可以通过改变模型参数调整检测出的欺诈行为的数目。就可视化而言,一个改进策略是边绑定与聚类,不过这可能会影响数据展示精度。

(2) 可视化设计选择。该系统主要是从“日期-参保人-医生-医院”的角度进行欺诈检测,在专家访谈过程中,他们指出药品的类别和单价也发挥着重要的作用,如何针对这些信息进行可视设计与编码,还需要进一步分析。

未来将考虑使用更细致的医保就诊数据用作

欺诈检测,以便探索出更有意义的欺诈模式.此外,可视化形式可以作为一种广泛的设计参考扩展已有的其他医保数据分析系统.最后,希望能够根据专家找到的可疑欺诈模式,自动地挖掘并推荐出其他相似的实例.

## 6.2 总结

本文介绍了 MedicareVis,以帮助医保审计人员识别和检查可疑的欺诈行为及其关联.该系统通过交互式分析处理流水线集成传统的数据挖掘算法与可视分析技术,并使用真实医保数据的实例研究以及对4位领域专家的访谈,展示了该系统的有效性和实用性.

## 参考文献(References):

- [1] Ren Y J, Shi Y L, Zhang K, *et al.* Medical treatment migration prediction based on GCN via medical insurance data[J]. *IEEE Journal of Biomedical and Health Informatics*, 2020, 24(9): 2516-2522
- [2] Demchenko Y, Grosso P, De Laat C, *et al.* Addressing big data issues in scientific data infrastructure[C] // *Proceedings of International Conference on Collaboration Technologies and Systems*. Los Alamitos: IEEE Computer Society Press, 2013: 48-55
- [3] Mackey T K, Liang B A. Combating healthcare corruption and fraud with improved global health governance[J]. *BMC International Health and Human Rights*, 2012, 12(1): Article No. 23
- [4] Zhang Y X, Chanana K, Dunne C. IDMVis: temporal event sequence visualization for type 1 diabetes treatment decision support[J]. *IEEE Transactions on Visualization and Computer Graphics*, 2019, 25(1): 512-522
- [5] Ni Mingxuan, Zhang Qian, Tan Haoyu, *et al.* Smart healthcare: from IoT to cloud computing[J]. *Scientia Sinica: Informationis*, 2013, 43(4): 515-528(in Chinese)  
(倪明选, 张黔, 谭浩宇, 等. 智慧医疗——从物联网到云计算[J]. *中国科学: 信息科学*, 2013, 43(4): 515-528)
- [6] Toyoda S, Niki N. Visualization-based medical expenditure analysis support system[C] // *Proceedings of the 37th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*. Los Alamitos: IEEE Computer Society Press, 2015: 1600-1603
- [7] Lin C H, Huang L C, Chou S C T, *et al.* Temporal event tracing on big healthcare data analytics[C] // *Proceedings of the IEEE International Congress on Big Data*. Los Alamitos: IEEE Computer Society Press, 2014: 281-287
- [8] Mohammed J, Lung C H, Ocleanu A, *et al.* Internet of things: remote patient monitoring using web services and cloud computing[C] // *Proceedings of the IEEE International Conference on Internet of Things, and IEEE Green Computing and Communications and IEEE Cyber, Physical and Social Computing*. Los Alamitos: IEEE Computer Society Press, 2014: 256-263
- [9] Bashyam V, Hsu W, Watt E, *et al.* Problem-centric organization and visualization of patient imaging and clinical data[J]. *RadioGraphics*, 2009, 29(2): 331-343
- [10] Ristevski B, Chen M. Big data analytics in medicine and healthcare[J]. *Journal of Integrative Bioinformatics*, 2018, 15(3): 20170030
- [11] Peng Yanni, Fan Xiaoping, Zhao Ying, *et al.* A survey of visualization of temporal event sequence data[J]. *Journal of Computer-Aided Design & Computer Graphics*, 2019, 31(10): 1698-1710(in Chinese)  
(彭燕妮, 樊晓平, 赵颖, 等. 时间事件序列数据可视化综述[J]. *计算机辅助设计与图形学学报*, 2019, 31(10): 1698-1710)
- [12] Ji Shouling, Li Jinfeng, Du Tianyu, *et al.* Survey on techniques, applications and security of machine learning interpretability[J]. *Journal of Computer Research and Development*, 2019, 56(10): 2071-2096(in Chinese)  
(纪守领, 李进锋, 杜天宇, 等. 机器学习模型可解释性方法、应用与安全研究综述[J]. *计算机研究与发展*, 2019, 56(10): 2071-2096)
- [13] Viveros M S, Nearhos J P, Rothman M J. Applying data mining techniques to a health insurance information system[C] // *Proceedings of the 22nd Very Large Data Base Conference*. Burlington: Morgan Kaufmann Publishers, 1996, 96: 286-294
- [14] Kumar M, Ghani R, Mei Z S. Data mining to predict and prevent errors in health insurance claims processing[C] // *Proceedings of the 16th ACM International Conference on Knowledge Discovery and Data Mining*. New York: ACM Press, 2010: 65-74
- [15] Skiba J M. A phenomenological study of the challenges and barriers facing insurance fraud investigators[D]. Minneapolis: Capella University, 2013
- [16] Ekina T, Leva F, Ruggeri F, *et al.* Application of Bayesian methods in detection of healthcare fraud[J]. *Chemical Engineering Transactions*, 2013, 33: 151-156
- [17] Johnson M E, Nagarur N. Multi-stage methodology to detect health insurance claim fraud[J]. *Health Care Management Science*, 2016, 19(3): 249-260
- [18] Shin H, Park H, Lee J, *et al.* A scoring model to detect abusive billing patterns in health insurance claims[J]. *Expert Systems with Applications*, 2012, 39(8): 7441-7450
- [19] Hubick K T. Artificial neural networks in Australia[M]. Canberra: CPN Publications, 1992
- [20] He H X, Wang J C, Graco W, *et al.* Application of neural networks to detection of medical fraud[J]. *Expert Systems with Applications*, 1997, 13(4): 329-336
- [21] Liou F M, Tang Y C, Chen J Y. Detecting hospital fraud and claim abuse through diabetic outpatient services[J]. *Health Care Management Science*, 2008, 11(4): 353-358
- [22] Li Z F, Liu X F, Cao X. A study on improved eclat data mining algorithm[J]. *Advanced Materials Research*, 2011, 328-330: 1896-1899
- [23] Agarwal R, Srikant R. Fast algorithms for mining association rules in large databases[C] // *Proceedings of the 20th International Conference on Very Large Data Bases*. Burlington: Morgan Kaufmann Publishers, 1994: 487-499
- [24] Han J W, Pei J, Yin Y W, *et al.* Mining frequent patterns without candidate generation[J]. *ACM SIGMOD Record*, 2000, 29(2): 1-12
- [25] He Z Y, Xu X F, Deng S C. Discovering cluster-based local outliers[J]. *Pattern Recognition Letters*, 2003, 24(9/10): 1641-1650