

Моделирование стационарного ведения пациентов на основе кластерного анализа

Мазелис Л.С.¹, Мазелис А.Л.¹, Смородин Г.Н.², Колесниченко О.Ю.³,
Мартынов А.В.⁴, Пулит В.В.⁴, Колесниченко Ю.Ю.³, Баландин С.И.⁵

¹ФГБОУ ВО «Владивостокский государственный университет экономики и сервиса», Владивосток

²Dell EMC External Research and Academic Alliances, Санкт-Петербург

³Uzgraph, Москва

⁴СП.АРМ, Санкт-Петербург

⁵Finnish-Russian University Cooperation in Telecommunications, Санкт-Петербург

В рамках совместного многоцентрового сотрудничества проводится исследование по анализу записей медицинских информационных систем (МИС). Были проанализированы записи МИС, сделанные в период с июня 2014 года по апрель 2015 года, с трех медицинских учреждений, оказывающих стационарную медицинскую помощь. Выборка больных для анализа составила 685 пациентов, все они проходили стационарное лечение с диагнозом по МКБ-10 – Код: I11.9: Гипертензивная (гипертоническая) болезнь с преимущественным поражением сердца без сердечной недостаточности; или Гипертоническая болезнь 2 стадии. Средний возраст пациентов составил 56,5 лет. Персональные данные пациентов не передавались, каждому пациенту был присвоен уникальный идентификатор.

Целью работы было выявление ряда различных взаимосвязей между характеристиками пациентов, что позволило бы сформировать более четкую картину исследуемой выборки пациентов и обоснованно принимать управленческие решения.

Каждому пациенту соответствовал вектор, включающий пять ключевых характеристик (длительность пребывания в стационаре, возраст, время ожидания предоставления последней услуги, стоимость лечения, общий результат лечения) и информацию об услугах, предоставленных пациенту в период прохождения лечения в стационаре. Информация по медуслугам состояла из следующего набора: для каждой услуги, из общего списка предоставляемых больницей, указывалось, сколько раз ее оказывали (число анализов, исследований, процедур) для конкретного пациента, и время ожидания каждой услуги.

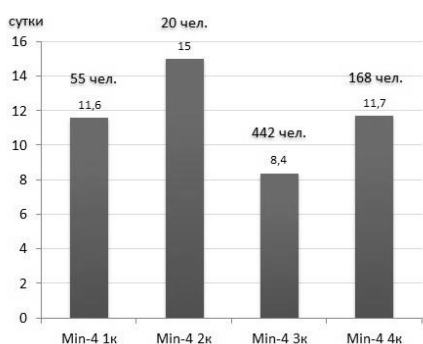
Набор медуслуг был разделен на 3 группы: минимальный набор (10 необходимых и наиболее часто реализуемых), укороченный (24 услуг, соответствующих Стандарту медпомощи при первичной артериальной гипертензии, приказ Минздрава РФ от 9 ноября 2012 г. №708н), максимальный (все услуги из 39 предоставленных). Анализировались 7 выборок: выборка ключевых характеристик $Q_{\text{ключ}}$ и выборки предоставленных услуг Q_{ik} , где i – номер группы услуг, $i=1,2,3$ и j – тип переменной ($j=1$ – количество предоставления каждой услуги, $j=2$ – время предоставления услуги).

В качестве инструмента для проведения анализа использован язык программирования Python. Для автоматизации расчетов была написана программа с использованием интерактивной среды iPython и библиотек NumPy, Pandas и Sklearn, предназначенных специально для анализа и обработки данных.

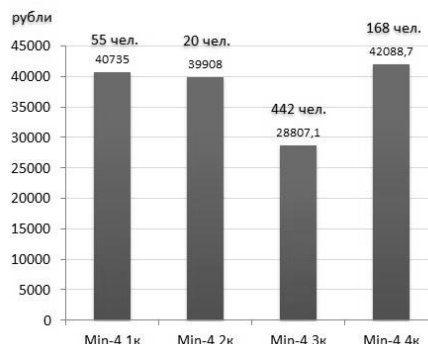
На первом этапе иерархическим методом и методом k-средних проведена кластеризация пациентов для описанных выборок. Иерархическая кластеризация позволила определиться с рациональным числом кластеров. Методом k-средних построены разбиения на 3, 4, 5 кластеров для каждой выборки. Для оценки валидности

разделения на кластеры использовали индекс Дэвиса-Болдина. Для каждого кластера вычислялись средние значения пяти ключевых характеристик, и проводилась проверка гипотез о статистической значимости различий выборочных средних этих характеристик для разных кластеров.

На втором этапе сделана количественная оценка степени совпадения кластеров, созданных для разных по типу и группе наборов данных. Анализ построенных матриц совпадения кластеров позволил обосновать гипотезу о наличии зависимости между принадлежностью пациента к определенному кластеру по минимальному набору данных и попаданием его в некий соответствующий кластер по максимальному набору, и оценить вероятность таких переходов. Выявление данной зависимости позволяет по принадлежности пациента к определенному кластеру по минимальному набору данных сделать прогноз по времени и стоимости лечения в стационаре (рис. 1).



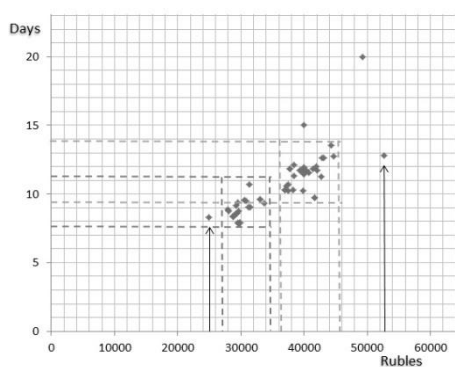
1.1



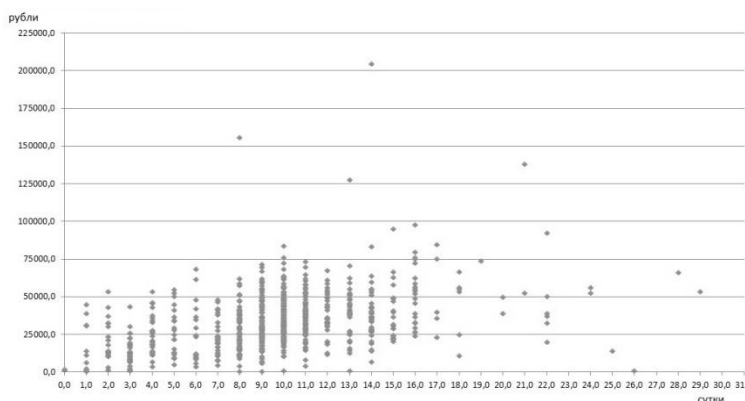
1.2

Рис. 1. Кластерное распределение по минимальному набору данных (средние значения для каждого кластера). 1.1 – длительность госпитализации. 1.2 – стоимость лечения. Обозначения по оси X – условные названия кластеров.

Проведена визуализация всех полученных кластеров на графике по критерию «стоимость лечения – длительность госпитализации» (средние показатели в каждом отдельном кластере), см. рис. 2.1. Для сравнения отражено простое распределение без кластеризации всех зарегистрированных случаев лечения (полная выборка 685 пациентов, см. рис. 2.2).



2.1



2.2

Рис. 2. 2.1 – Распределение всех кластеров по двум характеристикам (средние значения для каждого кластера) «стоимость лечения – длительность госпитализации». Стрелками указаны кластер с самой низкой стоимостью лечения и кластер с самой высокой стоимостью лечения. 2.2 – Полная выборка пациентов, простое распределение значений длительности госпитализации и стоимости лечения для каждого пациента.

Корреляционный анализ по всей выборке без кластеризации не показал сильных связей, а после кластеризации корреляционный анализ средних значений кластеров показал наличие сильной связи между длительностью госпитализации и стоимостью лечения (коэффициент корреляции Пирсона $r=0,8$). Также визуализация кластеров

позволила выявить распределение на две группы (группа 1 – меньшая длительность и стоимость стационарного лечения, группа 2 – большая длительность и стоимость), что стало основой для описания двух моделей ведения пациентов с гипертонической болезнью.

На рис. 3.1 показаны две обнаруженные модели стационарного ведения таких пациентов – с преимущественным обследованием сердечно-сосудистой системы (Модель А, группа 1, 58% пациентов) и преимущественным обследованием желудочно-кишечного тракта, сопутствующих заболеваний (Модель В, группа 2, 42% пациентов). На рис. 3.2 представлена визуализация корреляционного анализа некоторых средних для каждого кластера параметров ожидания с общим временем пребывания в стационаре. Выявлены сильные корреляционные связи, указывающие на то, что более всего влияет на длительность пребывания в стационаре ожидание взятия крови на анализ и ультразвукового обследования, а также сроки проведения парентеральной терапии.

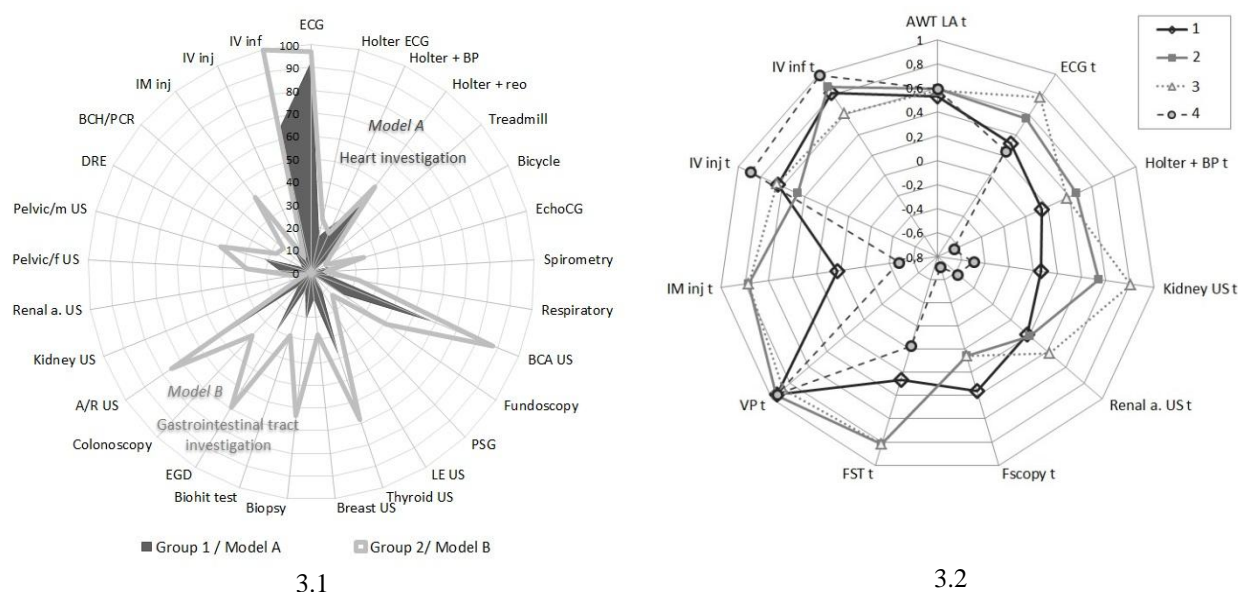


Рис. 3. 3.1 – Визуализация двух моделей стационарного обследования. Модель А: обследование сердечно-сосудистой системы (Model A, Heart investigation). Модель Б: обследование преимущественно желудочно-кишечного тракта, сопутствующих заболеваний (Model B, gastrointestinal tract investigation). Показано число обследований и процедур, стандартизированных на 100 пациентов. Сокращения и условные обозначения см. ниже. 3.2 – Корреляционный анализ по Пирсону между средним временем пребывания в стационаре и некоторыми средними параметрами ожидания процедур и обследований. 1 – все кластеры; 2 – 8 кластеров по максимальному набору медуслуг; 3 – 8 кластеров по укороченному набору; 4 – 8 кластеров по минимальному набору.

Сокращения для рис. 3.1: ECG – электрокардиография; Holter ECG – суточное мониторирование ЭКГ в 12-ти отведениях; Holter + BP – суточное мониторирование ЭКГ и артериального давления; Holter + reo – комбинированное суточное мониторирование ЭКГ в 12-ти отведениях, АД и реопульмограммы; Treadmill – тредмил стресс-тест; Bicycle – велоэргометрия; EchoCG – стресс-эхокардиография; Spirometry – спирометрия с пробой с бронхолитиком; Respiratory – комплексное исследование легких; BCA US – триплексное исследование брахиоцефальных артерий на интра- и экстракраниальном уровне; Fundoscopy – исследование глазного дна с применением линзы Гольдмана; PSG – комплексное полисомнологическое обследование; LE US – дуплексное исследование артерий и вен нижних конечностей, дуплексное исследование вен нижних конечностей с функциональными пробами; Thyroid US – УЗИ щитовидной железы; Breast US – УЗИ молочных желез; Biopsy – биопсия при эндоскопии; Biohit – тест Biohit; EGD – эзофагогастродуоденоскопия; Colonoscopy – колоноскопия; A/R US – УЗИ органов брюшной полости; Kidney US – УЗИ почек; Renal a. US – УЗИ почечной артерии; Pelvic/f US – УЗИ малого таза у женщин; Pelvic/m US – УЗИ малого таза у мужчин; DRE – ректальное пальцевое исследование предстательной железы; BCH/PCR – взятие материала на бактериологическое,

цитологическое, гормональное или ПЦР-исследование; *IM inj* – инъекция внутримышечная; *IV inj* – инъекция внутривенная; *IV inf* – внутривенное капельное введение лекарственных средств.

Сокращения для рис. 3.2: *AWT LA t* – среднее ожидание последнего анализа; *ECG t* – ожидание электрокардиографии; *Holter + BP t* – ожидание суточного мониторирования ЭКГ и артериального давления; *Kidney US t* – ожидание УЗИ почек; *Renal a. US t* – ожидание УЗИ почечной артерии; *Fscopy t* – ожидание исследования глазного дна с применением линзы Гольдмана; *FST t* – ожидание забора крови из пальца; *VP t* – ожидание забора крови из вены; *IM inj t* – ожидание внутримышечной инъекции; *IV inj t* – ожидание внутривенной инъекции; *IV inf t* – ожидание внутривенного капельного введения лекарственных средств.

Выявлен социально-медицинский портрет пациента (или iПациента), страдающего гипертензивной болезнью сердца: в основном такому пациенту требуется одно обследование ЭКГ; только в половине случаев проводится внутривенное капельное введение препаратов; почти равное внимание уделяется как обследованию сердца, так и обследованию желудочно-кишечного тракта, что оправдано с точки зрения оценки побочных эффектов антигипертензивных препаратов, а также указывает на наличие определенной направленности сопутствующих заболеваний у данной категории больных.

Анализ данных МИС показал, что 80-90% пациентов с диагнозом по МКБ-10 Код: П1.9 могут быть обследованы и пролечены в условиях дневного стационара без длительной госпитализации. Для укрепления такой тактики ведения пациентов необходимо развивать телемедицинские службы круглосуточного амбулаторного консультирования [1-3].

Список литературы:

1. Frolio L. Big Data Insights in Healthcare. Dell EMC Global Services Blog «In Focus». Hopkinton MA, USA, 2015.
2. Reys A., Balandin S. Healthcare, medical support and consultancy applications and services for mobile devices, in Proc. IEEE SIBIRCON Conf. July 2010, pp. 300-305.
3. Schmarzo B. Big Data MBA: Driving Business Strategies with Data Science. USA: John Wiley & Sons. Inc., 2016.