

BIG DATA: АНАЛИЗ ГОДОВОГО МАССИВА ВЫЗОВОВ НЕОТЛОЖНОЙ МЕДИЦИНСКОЙ ПОМОЩИ Г.БАРАНОВИЧИ, ИНСТРУМЕНТЫ И ТЕХНОЛОГИИ

С.Е. Полубинская, Е.С. Филипчик

Данная работа основана на информационной обработке годового массива исходных данных диспетчерской службы скорой медицинской помощи г. Барановичи. Описывается технология ректификации данного массива данных в среде пакета OpenRefine и последующий анализ временных рядов в статистической среде R. Выявлены временные зависимости в интенсивности вызовов и сделана попытка объяснения генеза найденных недельных ритмов интенсивности вызовов неотложной помощи.

Введение

Станция неотложной медицинской помощи это организация здравоохранения оказывающая помощь населению при угрожающих жизни состояниях, несчастных случаях, острых тяжелых заболеваниях и обострении хронических заболеваний. По своему определению эти состояния возникают внезапно и не могут быть запланированы, однако это не означает что работу по оказанию неотложной медицинской помощи нельзя или не нужно планировать.

Статистика вызовов скорой медицинской помощи является важнейшим источником информации для планирования и нормирования деятельности данной службы. Учет региональных особенностей в распределении вызовов позволяет выявить закономерности, на основании учета которых возможно научно-обоснованное распределение имеющихся в распоряжении станции неотложной медицинской помощи средств и ресурсов.

Важнейшие характеристики качества работы – «время доезда», «число вызовов в смену», «время на вызове» невозможно обеспечить без учета особенностей

Исходные данные

Согласно данным литературы [1] порядка 2/3 звонков в диспетчерскую службу неотложной скорой помощи приводят к выезду бригады. Мы также ограничимся в нашем исследовании анализом только таких звонков.

Диспетчерская служба станции неотложной скорой помощи г. Барановичи использует ведомственную информационную систему регистрации и обработки поступаемых вызовов. Однако следует отметить, что показатели, которые не используются далее в данной информационной системе для построения сводных отчетов не снабжены справочниками и допускают произвольный ввод со стороны оператора системы. Такая ситуация ведет к формированию массивов данных которые характеризуются термином “Messy Data”[2], и требует для использования таких данных значительных усилий по их предварительной обработке.

Данные представлены в формате таблиц выгруженных из ведомственной системы учета вызовов.

Таким образом стоят задачи:

1. Объединить данные журналов вызовов бригад скорой медицинской помощи и очистить их от опечаток и не точностей описания характеристик вызовов.
2. Построить сводные понедельные таблицы вызовов.
3. Выявить временные зависимости связанные с недельным циклом интенсивности вызовов скорой медпомощи.

Решить проблемы с неконсистентностью данных предназначена технология «ректификации данных», наиболее типичный представитель таких информационных систем является OpenRefine. Это `www` приложение построенное на технологиях `java` (`Apache-Spark`, `mllib` и т. п.). Особенность приложения является архитектура системы с «бесконечной историей» изменений сделанных пользователем, и комплементарной возможностью делать сохранение-применение части записанной истории изменений одного набора данных к новым данным.

Для подготовки данных журнала вызовов основной функций OpenRefine была кластеризация значений текстовых полей записей. Встречается до 25 вариантов написания одних и тех же по смыслу меток. Подготовленные нормализованные значения текстовых полей можно использовать в качестве справочников в основном приложении журнала регистрации вызовов.

На основании поля даты регистрации с использованием библиотеки `lubridate` [4] реализующем модель календаря построены дополнительные признаки:

Анализ данных

После группировки данных с помощью пакета `reshape2` [5] получены итоговые почасовые и понедельные таблицы частоты вызовов на текущий интервал группирования. Наиболее удобно представить суточный цикл колебания частоты вызовов в зависимости от дня недели (см. рис. 1).

Немного выделяется воскресенье большим числом вызовов в дневное время, но в целом профиль вызовов в течении дня стабилен. Максимум в 22-23 часа, второй максимум в 12-13 часов. Минимумы наблюдаются в 5 часов и в 17 часов.

Однако наибольший интерес представляет выраженная неоднородность распределения вызовов внутри недельного интервала учета.

Прибегнем к анализу по методу принципиальных компонент. Картина корреляций показывает наличие зависимости “тяжелое начало недели - легкий конец недели”. Однако практически вся наблюдаемая картина обусловлена наличием неполных недель с выпадающей историей регистрации вызовов. Для более глубокого анализа исключим такие случаи (23, 22, 32, 13, 14).

Картина корреляционных связей показывает противопоставление выходных дней и понедельника остальным дням недели.

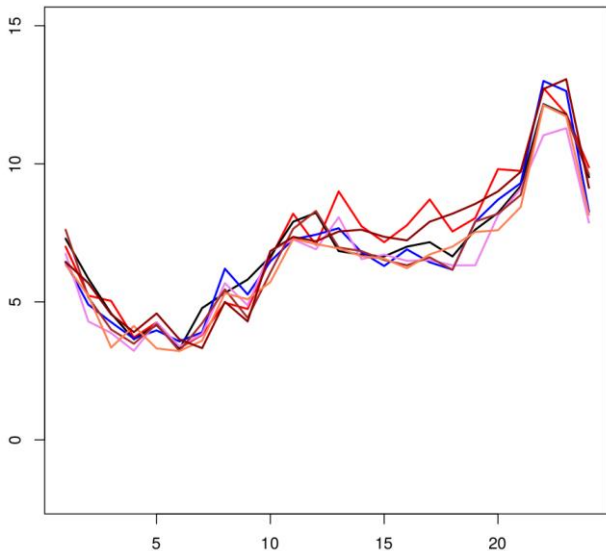


Рис.1. Внутрисуточная интенсивность поступления заявок (по горизонтали часы суток, по вертикали число вызовов)

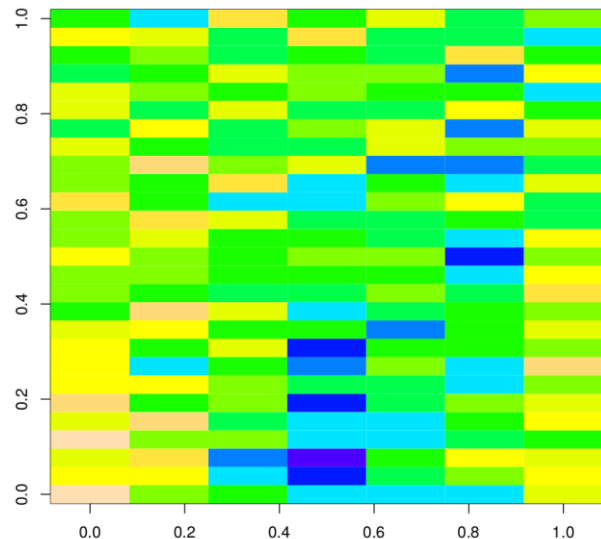


Рис.2. Внутринедельное распределение интенсивность поступления заявок (по горизонтали дни недели, по вертикали реализации недель отсортированные по величине первой принципиальной компоненты)

Нагруженное воскресенье, суббота и понедельник дает в результате разгруженную от вызовов среду (нижняя часть цветной матрицы). Равномерно нагруженное начало недели часто дает минимум нагрузки в пятницу (см. рис. 2, верхние строки цветной матрицы).

Заключение

Можно сделать вывод, что частота вызовов скорой медицинской помощи имеет выраженную суточную и внутри-недельную зависимость, позволяющую учесть дополнительные параметры в планировании организации работы скорой медицинской помощи.

Интенсивность вызовов скорой помощи зависит от факторов связанных с недельным циклом жизни социума и особенностей функционирования службы скорой медпомощи. Знание точных статистических характеристик этих процессов позволяет выявить резервы обслуживания скорой медицинской помощи и ведущие факторы лимитирующие эти возможности.

Список литературы

1. Миняев В.А., Вишняков Н.И., Юрьев В.К., Лучкевич В.С. Социальная медицина и организация здравоохранения. — С.-Петербург, 1998. Издательство “Водолей”, II том. — 439 с.
2. Johnson, D. E. and Milliken, G. A. 2006. Messy Data. Encyclopedia of Statistical Sciences. 7.
3. Huynh D., Mazzocchi S. OpenRefine. – 2012.

4. Garrett Grolmund, Hadley Wickham (2011). Dates and Times Made Easy with lubridate. Journal of Statistical Software, 40(3), 1-25. URL <http://www.jstatsoft.org/v40/i03/>.
5. Hadley Wickham (2007). Reshaping Data with the reshape Package. Journal of Statistical Software, 21(12), 1-20. URL <http://www.jstatsoft.org/v21/i12/>.

Полубинская Светлана Евгеньевна, старший преподаватель кафедры спортивной медицины и ЛФК Гродненского государственного медицинского университета, sveta.polubinskaja@gmail.com

Филипчик Евгений Сергеевич, студент 4 курса специальности 1-01 02-03 «Физическая культура» факультета физической культуры Гродненского государственного университета имени Янки Купалы, vip.filipchik@mail.ru