

С.З. УМАРОВ, д.фарм.н., профессор

К.И. НАРКЕВИЧ, Санкт-Петербургская государственная химико-фармацевтическая академия Минздрава России

10.21518/1561-5936-2018-1-2-60-63

# Анализ интенсивности посетительского трафика в фармацевтическом ретейле с использованием элементов обработки больших данных (BIG DATA)

Извлечение ценной информации из огромных объемов данных становится ключевым фактором конкуренции и развития компаний и создает предпосылки для изменения бизнес-процессов, действующих в фармацевтической отрасли. Одним из методов обработки больших данных является анализ видеоданных, полученных из различных источников.

Провести анализ значительных объемов видеoinформации могут системы интеллектуального видеонаблюдения, которые являются комплексом аппаратных и программных средств, предназначенных для автоматического обнаружения событий, определенных набором заранее заданных критериев, и реакции на их обнаружение по установленному правилу в режиме реального времени. В работе использовалась современная система видеомониторинга с уникальным комплексом видеоаналитики, обладающая функцией модуля «Аналитика подсчета посетителей». В результате анализа 800 ГБ видеоданных была получена информация о динамике посетительского трафика в различные периоды рабочего дня и установлено, что среднее число посетителей в единицу времени (час) зависит как от времени года, так и от времени суток. Тем не менее вариабельность посетительского трафика аптеки нельзя назвать высокой, ее можно охарактеризовать как умеренную (минимум 17%, максимум 40%). Но наличие даже умеренной вариабельности посетительского трафика сказывается на интенсивности труда фармацевтического персонала.

## Ключевые слова:

большие данные, фармацевтический ретейл, видеоаналитика, анализ посетительского трафика

Современная экономика все больше и больше завоевывает цифровое пространство. Впечатляющее снижение затрат, требуемых на получение, обработку и передачу информации, изменяет способ ведения бизнеса. Виртуальное пространство, в свою очередь, оказывает все более интенсивное воздействие на человека и его деятельность. Это объясняется тем, что кардинальные инновации в вычислительной технике и методах создания, передачи и обработки информации всех видов приводят к зарождению принципиально новых культурных тенденций

## SUMMARY

**Keywords:** big data, pharmaceutical retail, video analytics, visitor traffic analysis

Extracting valuable insights from vast quantities of data becomes a key factor in developing new competitive advantages and companies' growth and creates the prerequisites for changing business processes in the pharmaceutical industry. The analysis of video data obtained from various sources is one of the methods for processing big data. Intelligent video surveillance systems built on hardware devices enabling solutions designed to automatically detect events that will match the predefined criteria and reaction to the systems disclosure according to the established rule in real time can analyse significant volumes of video information. The analysis was performed using a state-of-the-art video monitoring system with a unique video analytics complex, which has the function of the People Counting module. The analysis of 800 GB of video data provided information on the visitor traffic dynamics during various working hours, which showed that the average number of visitors per time unit (hour) depends both on the time of year and the time of day. Nevertheless, the variability of pharmacy visitor traffic cannot be called high; it can be characterized as moderate (minimum 17%, maximum 40%). But even a moderate variability of visitor traffic affects the intensity of labour of pharmaceutical personnel.

S.Z. UMAROV, Pharm.D, Prof., K.I. NARKEVICH, Saint Petersburg State Chemical Pharmaceutical Academy of the Ministry of Health of Russia. PHARMACEUTICAL RETAIL VISITOR TRAFFIC INTENSITY ANALYSIS USING BIG DATA TECHNOLOGY ELEMENTS

в жизни общества и восприятии окружающего мира, причем это влияние носит глобальный характер благодаря не только успешному развитию новых технологий, но и их широкому использованию [1].

В информационном обществе в значительной степени изменяются представления о роли и возможностях использования информационных ресурсов в различных сферах жизни. В связи с доступностью и простотой распространения информации новые технологии по работе с ней находят широкое применение в политической, экономической и социальной жизни. Информация используется, прежде всего, как средство целевого массового воздействия на состояние общества, становясь механизмом управления им, на потенциал развития человека [2]. Актуальная и достоверная информация, предоставляемая в формате открытых данных, становится двигателем развития разных областей.

Вышеизложенное относится и к сфере фармацевтической деятельности, когда перед специалистом стоит задача наблюдения (контроля) за объектами, явлениями и процессами фармацевтического рынка. В результате наблюдения происходит получение информации в информационном поле [3, 4], накопление опыта и формирование описаний объектов, явлений и процессов.

Рост объемов собираемой информации и требование ее обработки и хранения делают актуальными исследования в области методов и алгоритмов анализа больших и сверхбольших массивов данных.

Впервые термин «большие данные» был предложен в 1998 г. Джоном Мэши, главным специалистом компании Silicon Graphics, но в то время термин не стал популярным [5]. Проблемой больших данных в нашей стране занимался академик В.М. Глушков, который еще в 1987 г. охарактеризовал Big Data как «информационный барьер» [6].

Одними из первых предложили использовать большие данные в практических целях специалисты из исследовательско-консалтинговой компании Gartner. Ими было предложено определение Big Data

*«...как набора данных, сбор, управление и обработку которых невозможно осуществить с помощью наиболее часто используемых аппаратных сред и программных инструментов в течение допустимого для пользователя времени» [7].*

Начиная с 2009 г. Big Data стали интересовать не только ученые, но и бизнесмены различных отраслей, включая фармацевтическую. В 2010 г. появились первые специальные программные и технические продукты, предназначенные исключительно для обработки больших объемов данных. С 2011 г. крупнейшие поставщики информационных технологий IBM, Oracle, Microsoft, Hewlett-Packard, EMC стали ориентировать стратегии своего развития на большие данные.

К 2020 г. прогнозируется увеличение объемов информации более чем в 50 раз, что, в свою очередь, потребует приращения количества необходимых для ее хранения серверов в 10 раз. Это, безусловно, сулит более чем радужные перспективы для производителей не только различного рода носителей информации, но и всех соответствующих деталей и комплектующих, кабельной продукции, прочего сырья для модулей [8].

Однако гораздо более существенный эффект обеспечивается не столько хранением, сколько обработкой и анализом информации, что позволяет извлекать экономическую выгоду именно из больших объемов данных, многообразия их видов, источников, сфер, а также высокой скорости обновления и пополнения, причем круг бенефициаров не ограничивается исключительно информационно-коммуникационными компаниями (ИКТ-компаниями) [9]. Вышеизложенное также может представлять интерес и для фармацевтической отрасли.

Для России тема Big Data актуальна ничуть не меньше, чем для остального мира. По разным оценкам, цифровая экономика составляет от 3 до 21% ВВП стран «Большой двадцатки», однако отечественная экономика еще не добилась значимых результатов в работе с большими данными. Сегодня в России подобными технологиями интересуются в основном

коммерческие предприятия: торговые сети, банки, телекоммуникационные компании. По оценке Российской ассоциации электронных коммуникаций, объем цифровой экономики в РФ в 2015 г. составил всего 1 трлн руб. (около 1,5% от ВВП). Тем не менее у нашей страны есть огромный потенциал роста цифровой экономики [10]. Так, согласно результатам нового исследования CNews Analytics и Oracle, уровень зрелости российского рынка Big Data за последний год повысился. Респонденты, представляющие 108 крупных предприятий из разных отраслей, продемонстрировали более высокую степень осведомленности об этих технологиях, а также сложившееся понимание потенциала подобных решений для своего бизнеса. Готовность к проектам Big Data растет в разных отраслях экономики России. Более трети опрошенных компаний (37%) приступили к работе с технологиями Big Data, из них 20% уже используют такие решения, а 17% начинают экспериментировать с ними. Еще более трети (36%) респондентов в настоящий момент рассматривают возможность их использования в ближайшей перспективе [11].

В «Стратегии развития отрасли информационных технологий в Российской Федерации на 2014–2020 годы и на перспективу до 2025 года» технологии обработки больших данных обозначены в числе «прорывных для мировой индустрии, в которых в перспективе 10–15 лет с высокой вероятностью может быть обеспечена глобальная технологическая конкурентоспособность России» [12].

Перед большинством фармацевтических фирм России стоит цель максимально использовать имеющиеся корпоративные ресурсы для обеспечения стабильного развития. Стратегию бизнеса следует формировать на основе оценки ресурсов самого предприятия и состояния внешней среды [13].

Появление новых данных и методов анализа создает предпосылки для изменения бизнес-процессов в компаниях, работающих в фармацевтической отрасли. Особенно это актуально для фармацевтического ретейла, где в прошлом ни данные, ни аналитика

ТАБЛИЦА &gt; Показатели интенсивности посетительского трафика в аптеке

Период наблюдения		Число наблюдений (час)	Среднее значение (чел.)	Минимальное значение (чел.)	Максимальное значение (чел.)	Стандартное отклонение (чел.)	Коэффициент вариации (%)
Время года	Время суток						
ЗИМА	Утро	200	13	7	19	2	19
	День	200	16	7	28	4	24
	Вечер	200	18	7	34	4	25
ВЕСНА	Утро	200	15	7	22	2	17
	День	200	17	5	25	3	19
	Вечер	200	19	7	33	5	24
ЛЕТО	Утро	200	12	4	19	2	20
	День	200	14	5	26	3	24
	Вечер	200	15	3	29	5	31
ОСЕНЬ	Утро	200	16	6	26	3	20
	День	200	18	5	28	4	24
	Вечер	200	15	0	34	6	40

не играли большой роли. Качество и актуальность данных имеют определяющее значение для результатов анализа.

Конечно, с точки зрения бизнеса большие данные – это всего лишь технология, рассматриваемая как «черный ящик». Она имеет свои преимущества и недостатки и позволяет, попав к грамотному руководителю, повысить прибыль предприятия.

Краткий список методов анализа структурированных и неструктурированных данных в технологии Big Data Analysis включает [14]:

1. Методы интеллектуального анализа Data Mining, а именно: обучение ассоциативным правилам, классификацию, кластерный анализ и регрессионный анализ.
2. Методы прогнозной аналитики.
3. Статистический анализ.
4. Визуализацию аналитических данных.

В соответствии с вышесказанным нами был предложен алгоритм динамической оценки трафика посетителей на основе существующих систем видеонаблюдения. Интеллектуальное видеонаблюдение (IVS) или видеоаналитика (video analytics) анализируют видео без прямого участия человека, опираясь на алгоритмы обработки изображения и распознавания образов.

Видеоаналитика используется в составе интеллектуальных систем видеонаблюдения. Системы интеллектуаль-

ного видеонаблюдения – это комплекс аппаратных и программных средств, предназначенных для автоматического обнаружения событий, определенных набором заранее заданных критериев, и для реакции на обнаружение по установленному правилу в режиме реального времени.

В настоящее время широкое применение нашли такие системы интеллектуального видеонаблюдения, как Masgoscor и ITV, 3VR. Интеллектуальное видеонаблюдение Masgoscor способствует увеличению функциональности системы видеонаблюдения и включает следующие интеллектуальные модули: модуль интерактивного поиска, контроль кассовых операций; модуль трекинга, позволяющий отслеживать движущиеся объекты в поле зрения камеры; модуль подсчета посетителей; модуль обнаружения и распознавания лиц; модуль распознавания номеров.

Интеллектуальное видеонаблюдение ITV предназначено для различных областей человеческой деятельности, инновационные разработки ITV включают видеоаналитику программного комплекса «Интеллект» (система видеодетекторов и система интеллектуального поиска в видеоархиве), модуль распознавания автомобильных номеров (распознавание государственных регистрационных знаков транспортных средств по видеоизображению) и др. Интеллектуальное видеонаблюдение 3VR – современная система видео-

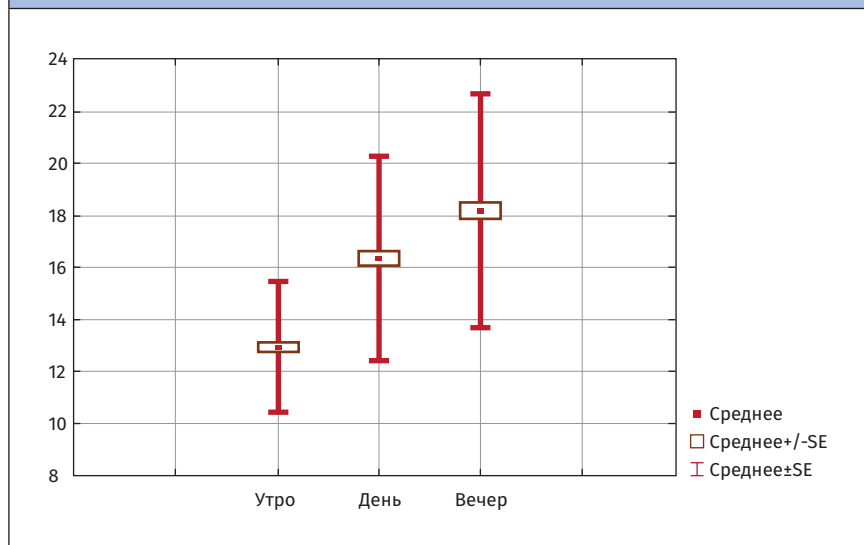
мониторинга с уникальным комплексом видеоаналитики, применение которой позволяет не только фиксировать события, но и предупреждать возможные угрозы оперативным реагированием на них [15]. Нами было предложено использовать функции модуля «Аналитика подсчета посетителей», что позволило определять количество вошедших и вышедших из аптеки людей в реальном времени, а также строить отчеты о находящихся в торговом зале посетителей за любой промежуток времени по одной или нескольким камерам. В результате были получены данные о динамике посетительского трафика в различные периоды рабочего дня: утро (8:00–12:00), обед (13:00–17:00) и вечер (18:00–21:00). Исходный объем видеoinформации, подлежащий первичной обработке, составил более 800 Гб. Однако после первичной обработки полученные данные стало возможным выгрузить сначала в стандартный табличный процессор Excel, а затем провести обработку. Структура данных, предназначенных для дальнейшей обработки, представляла собой набор из 12 переменных и 200 случаев наблюдений. Каждая из переменных представляла собой сведения о динамике посетительского трафика в конкретном временном срезе (утро, день, вечер) за каждый час. Количество наблюдений для каждой переменной составило 200 часов, что позволило

охватить в целом 2 400 рабочих часов, которые составляют 66,7% всего годового бюджета рабочего времени (при 10-часовом рабочем дне). Результаты статистической обработки данных и полученные на их основе показатели интенсивности посетительского трафика в аптеке представлены в *таблице*.

Как следует из данных, представленных в таблице, среднее число посетителей в единицу времени (час) зависит как от времени года, так и от времени суток. Тем не менее вариабельность посетительского трафика аптеки нельзя назвать высокой, скорее, ее можно охарактеризовать как умеренную (минимум 17%, максимум 40%). Однако наличие даже умеренной вариабельности посетительского трафика сказывается на интенсивности труда фармацевтического персонала, рабочая смена которого представлена двумя специалистами.

Для объективизации утверждения о достоверности различия показателей посетительского трафика в различное время суток был проведен сравнительный анализ средних значений посетительского трафика в зимний период в различное время суток. С этой целью нами был применен метод анализа нескольких выборок. Важнейшим вопросом, возникающим при анализе выборок, является вопрос о наличии различий между ними. В каче-

РИСУНОК ► Динамика посетительского трафика в зимнее время



стве меры оценки был использован критерий Стьюдента (t-критерий), который позволяет найти вероятность того, что рассматриваемые средние значения в выборке относятся к одной и той же совокупности. Данный критерий наиболее часто используется для проверки гипотезы: «средние двух выборок относятся к одной и той же совокупности» (нулевая гипотеза). Соответственно, альтернативная гипотеза основана на противоположном суждении и предполагает существенные различия между средними величинами. В результате анализа посетительского трафика в зимний период

в различное время суток было установлено достоверно статистически значимое различие, нашедшее графическое отражение на диаграмме размаха (*рис.*).

Таким образом, на основе использования элементов анализа Big Date стало возможным сделать первые попытки прогнозирования социально-демографических и поведенческих характеристик посетителей и создания профиля «типичного клиента». Это может быть полезным как для оперативного управления, так и для таргетирования рекламы, анализа рекламных каналов и др.

#### ИСТОЧНИКИ

1. Портер М.Э. Конкуренция. Пер. с англ. М.: Издательский дом Вильямс, 2005. 608 с.
2. Булгакова Е.В., Булгаков В.Г., Акимов В.С. Использование больших данных в системе государственного управления: условия, возможности, перспективы. Юридическая наука и практика: вестник Нижегородской академии МВД России, 2015, 3(31): 10–14.
3. Денисов А.А. Информационное поле. М.: «Омега», 1998. 64 с.
4. Цветков В.Я. Естественное и искусственное информационное поле. Международный журнал прикладных фундаментальных исследований, 2014, 5, ч. 2: 178–180.
5. Большие Данные – новая теория и практика [Электронный ресурс]. Открытые системы. Режим доступа <https://www.osp.ru/os/2011/10/13010990>.
6. Глушков В.М. Основы безбумажной информатики. М.: Наука, 1987. 557 с.
7. Адриан М. Большие данные. Teradata Magazine, 2011, 11: 38–42.
8. Васёв П.А. Среда поддержки интерактивной визуализации для суперкомпьютерных вычислений. Вопросы атомной науки и техники. Серия: Математическое моделирование физических процессов, 2009, 4: 67–77.
9. Авербух В.Л., Байдалин А.Ю., Васёв П.А. и др. Задачи визуализации параллельных вычислений. Вопросы атомной науки и техники. Сер.

10. Математическое моделирование физических процессов, 2002, 3: 40–52.
11. Короткова Т. EMC Data Lake 2.0 – средство перехода к аналитике больших данных и цифровой экономике [Электронный ресурс]. CNews – издание о высоких технологиях. Режим доступа [http://bigdata.cnews.ru/news/line/2015-12-03\\_emc\\_data\\_lake\\_20\\_pomozhet\\_perejti\\_k\\_analitike](http://bigdata.cnews.ru/news/line/2015-12-03_emc_data_lake_20_pomozhet_perejti_k_analitike).
12. Исследование Oracle и CNews Analytics: Большие данные пришли в Россию [Электронный ресурс]. Оракл Россия и СНГ. Режим доступа. <https://www.oracle.com/ru/corporate/pressrelease/study-of-oracle-and-cnews-analytics-20150226.html>.
13. Распоряжение Правительства РФ от 01.11.2013 N 2036-р «Об утверждении Стратегии развития отрасли информационных технологий в Российской Федерации на 2014 – 2020 годы и на перспективу до 2025 года» [Электронный ресурс]. Система Гарант. – Режим доступа <http://base.garant.ru/70498122/#ixzz513wEv1HV>.
14. Головина Т.А., Романчин В.И., Закиров А.И. Развитие технологий бизнес-аналитики на основе концепции Business Intelligence. Известия Тульского государственного университета. Экономические и юридические науки, 2014, 5–1: 416–424.
15. Tan P, Steinbach M, Kumar V. Introduction to Data Mining. N.Y.: Addison-Wesley, 2005. 769 p.
16. Интеллектуальная видеоаналитика [Электронный ресурс]. Авнор – технические средства охраны. Режим доступа: [http://aviorst.ru/services/intellektualnaya\\_videoanalitika/](http://aviorst.ru/services/intellektualnaya_videoanalitika/).