

РЕКОМЕНДАТЕЛЬНЫЕ СИСТЕМЫ В ЗДРАВООХРАНЕНИИ

Цурко В. В.¹

(ФГБУН Институт проблем управления
им. В.А. Трапезникова РАН, Москва)

Рекомендательные системы получили широкое распространение в различных интернет-сервисах: контекстная реклама, интересные для пользователя фильмы и музыка, подбор товаров и услуг в интернет-магазинах. Алгоритмы построения рекомендаций учитывают профиль пользователя, историю его запросов, покупок и информацию о пользователях, похожих на него. Эти алгоритмы быстро развиваются и доказали свою эффективность, их внедрение повышает посещаемость сервисов и продажи. Новой областью применения рекомендательных систем является здравоохранение. За десятилетия накоплены огромные объемы медицинских данных, таких как истории болезни, базы смертности и заболеваемости, протоколы лечения и даже данные о здоровье непрерывно снимаемые с человека специальными датчиками. Разрабатываются интеллектуальные системы диагностики заболеваний, выбора способа лечения, отслеживания ежедневного здоровья человека. В статье приводится общее введение в рекомендательные системы, обзор литературы о применении рекомендательных систем в здравоохранении, а также рассматривается проблема построения рекомендательных систем для оптимизации больничных закупок.

Ключевые слова: рекомендательные системы, здравоохранение, оптимизация больничных закупок.

1. Введение

По мере накопления данных по поисковым запросам в Интернете, по результатам выбора предлагаемых услуг и товаров появилась возможность создавать индивидуальные и групповые профили пользователей, на основании которых разработаны компьютерные системы для прогноза товаров и услуг, которые могут заинтересовать конкретного пользователя в будущем. Такие системы получили название *рекомендательные*.

Рекомендательные системы широко распространены в Интернете, помогают людям найти релевантный товар, а компани-

¹ Варвара Владимировна Цурко, к.т.н., с.н.с. (v.tsurko@gmail.com).

ям – повысить прибыль. Алгоритмы учитывают профиль пользователя, историю его запросов, покупок и информацию о пользователях, похожих на него, для того чтобы предсказать, какая музыка, ресторан, отель или товар в интернет-магазине могут ему понравиться.

Методология создания рекомендательных систем является универсальным методом машинного обучения и может быть распространена на другие сферы, не связанные с интернетом. Это, например, финансы, управление дорожным движением, распределение ресурсов, оптимизация подготовки спортсменов к соревнованиям. Перспективным является внедрение рекомендательных систем в здравоохранении. Как часть системы интеллектуальной поддержки врача, такая система может применяться не только на этапе диагностики заболевания и выбора оптимального лечения, но и на этапе планирования закупок лекарств и оборудования, формирования кадрового состава специалистов для более полного удовлетворения нужд конкретного лечебного учреждения.

Статья посвящена обзору публикаций по применению рекомендательных систем в здравоохранении и адаптации методов построения рекомендательных систем для целей оптимизации больничных закупок. В разделе 2 вводится формализация понятия «рекомендательная система», примеры рекомендательных систем и алгоритмы их построения. В разделе 3 описаны применения рекомендательных систем в здравоохранении, адресованные врачам и пациентам. Врачам такие системы помогают точнее ставить диагноз, а пациентам для снижения индивидуальных рисков развития заболевания такие системы могут предоставлять рекомендации, например, по правильному питанию и здоровому образу жизни. В разделе 4 рассматривается формализация проблемы и методы построения рекомендательных систем для закупки лекарств для больниц.

2. Рекомендательные системы

Для лучшего понимания контекста последующих частей статьи приведем основные термины теории рекомендательных систем.

2.1. Определения и обозначения

Пользователь/субъект – человек, использующий систему и получающий рекомендации.

Объект (предмет, товар и т.п.) – объект, который рекомендуют пользователю.

Рейтинг – рейтинг проставляется пользователем на объект/товар. Рейтинг может проставляться явно или неявно, т.е. зависеть от активности пользователя (перешел по ссылке, открыл документ). Рейтинг характеризует ценность объекта для пользователя.

Матрица рейтингов – матрица, по строкам которой находится информация о всех субъектах/пользователях, а по столбцам – о всех возможных объектах. Элементами матрицы является информация о рейтинге конкретного объекта для пользователя. Если пользователь еще не оценил объект, то соответствующее значение в матрице пусто. Часто матрица бывает разреженной, т.е. большинство ее элементов пусто.

Проблема холодного старта – ситуация, при которой для новых пользователей или объектов система содержит мало информации для прогнозирования достоверного рейтинга объекта или построения надежной рекомендации субъекту.

2.2. Примеры рекомендательных систем

Цель рекомендательной системы заключается в нахождении объектов (фильмов, товаров, отелей и т.п.), релевантных для пользователя, на основании его предыдущих решений, пользовательского поведения.

Наиболее распространены рекомендательные системы, советующие купить тот или иной товар. Субъектом является пользователь интернет-магазина, объектом – товары. Рейтинг формиру-

ется на основе информации о покупках пользователей. Задачами подобной рекомендательной системы могут являться не только рекомендации наиболее релевантного товара клиенту, но и предложение взаимодополняемых товаров, информирование о новых товарах. Классическим примером рекомендательной системы товаров являются рекомендации в интернет-магазине Amazon.

Существуют рекомендательные системы, дающие советы пользователям, какую веб-страницу посетить. В этом случае субъектами являются пользователи сети Интернет, объектами – веб-страницы. Рейтинги проставляются на основе истории посещений пользователей. Задача – выбрать для пользователя ранжированный список интересных ему ресурсов. Примеры таких систем: Яндекс.Дзен, Surfingbird.

В 2006 году компания Netflix предоставила открытую выборку пользовательских рейтингов фильмов и предложила всем желающим поучаствовать в соревновании предсказания рейтингов оставшейся, неопубликованной части данных. Приз составлял один миллион долларов. Соревнование привело к быстрому развитию алгоритмов построения рекомендательных систем и созданию эффективного алгоритма построения рекомендаций – Bellkor's Pragmatic Chaos [8].

2.3. Алгоритмы

Центральным аспектом при построении рекомендательной системы является выбор алгоритма. Использование того или иного алгоритма обусловлено исходными данными, их размерностью и контекстом рекомендаций. В данной части будут описаны основные алгоритмические подходы к построению рекомендательных систем.

Метод *коллаборативной фильтрации* основан на индивидуальных рейтингах пользователей. Метод пытается предсказать наиболее релевантный объект для пользователя, используя матрицу рейтингов. Обычно матрица рейтингов является разреженной. Существуют различные способы заполнения пустых значений в матрице, которые могут быть использованы в качестве рекомендаций.

Простейшими являются *корреляционные модели* коллаборативной фильтрации (Memory-based Collaborative Filtering). Их особенностями являются хранение всей матрицы рейтингов и измерение сходства между объектами и пользователями с помощью корреляции строк и столбцов матрицы. Для заполнения пропусков в данных используется непараметрическая регрессия для пользователей или для объектов. Корреляционные методы интуитивно понятны, легко реализуемы. Основными недостатками корреляционных методов являются проблема холодного старта и необходимость хранения огромной разреженной матрицы рейтингов. Латентные модели лишены этих недостатков.

Для увеличения производительности методов используются *латентные модели* коллаборативной фильтрации. По матрице рейтингов оценивается профиль пользователей и профиль объектов, имеющие существенно меньшую размерность, чем исходные данные. Латентные модели могут быть основаны на кластеризации, вероятностном подходе и матричных разложениях. Чаще всего используется сингулярное разложение матриц.

Для оценки качества рекомендательной системы обычно используются метрики из теории информации: точность, чувствительность, F-мера, ROC-кривая, RMSE [9].

3. Рекомендательные системы в здравоохранении

Рекомендательные системы в здравоохранении не получили пока столь широкого распространения, как в интернет-бизнесе и торговле. Тем не менее, они имеют большой потенциал. В первую очередь, благодаря большому объему медицинских данных, протоколов лечения и тому факту, что многие процессы и принятие решений в области здравоохранения возможно автоматизировать с помощью применения методов машинного обучения.

Цифровые технологии и рекомендации могут сопровождать человека на разных этапах его жизни и лечения. Прежде всего следить за здоровым образом жизни: отслеживать необходимую физическую активность, время сна, питание, помогать при отказе от вредных привычек, контролировать уровень стресса и сохра-

нять концентрацию в течение дня [3]. Основанием для рекомендаций по здоровому образу жизни является заполненная пользователем информация о физическом состоянии, хронических заболеваниях, занесенный в программу рацион питания и активность человека в течение дня. Последняя может быть измерена специальным носимым оборудованием, таким как «умные» браслеты и часы, или программным обеспечением, установленным на смартфон. Человек в этом случае получает рекомендации на основе личной статистики и общепризнанных медицинских рекомендаций по поддержанию здоровья.

Автоматизированные системы принятия решений могут выявлять случаи отклонения от нормы и симптомы заболеваний, выдавать рекомендации о посещении врача или сдачи анализов. Для определения симптомов заболеваний используют специальные биосенсоры, такие как датчики давления, измерения глюкозы в крови, пульсометры и термометры. Хотя последние тенденции направлены на то, чтобы встраивать биосенсоры в повседневные вещи: мобильные телефоны или часы. Так, Apple Watch могут измерять частоту сердцебиения, снимать электрокардиограмму и сообщать своему владельцу случаи отклонения от нормы. Встроенные в часы акселерометр и гироскоп определяют падение человека, а устройство вызывает помощь, если человек не отвечает более 60 секунд [2].

Методы машинного обучения могут помочь при постановке диагноза. Имея медицинскую карту пациента, методами коллаборативной фильтрации можно показать, какие заболевания чаще всего встречаются у людей, имеющих похожую историю болезни, и, таким образом, предсказать заболевания, которые могут быть пока не диагностированы у пациента. Распознавание образов уже помогает выявлять аутизм по видеосъемке ребенка, что позволяет диагностировать заболевание на ранней стадии и достичь лучших результатов в лечении [4].

Актуально развитие рекомендательных систем, дающих советы пациентам, нуждающимся в постоянном приеме лекарств и отслеживании показателей здоровья. Больные сахарным диа-

бетом могут заносить данные о чувстве голода, болях, уровне глюкозы и приеме лекарств в мобильное приложение, которое агрегирует все показатели и напоминает о приеме необходимых лекарств [6]. Аналогично программа EpiWatch [7] позволяет больным эпилепсией вести учет возникновения и длительности приступа и регистрировать принимаемые медикаменты. Строя корреляционную модель между приемом лекарств и приступами, программа, установленная на смартфон, прогнозирует возможный приступ и при необходимости может вызвать помощь.

В статье [11] обсуждаются два основных сценария построения рекомендательных систем в здравоохранении. Первый сценарий рассматривает пациентов как конечных пользователей; он имеет огромный потенциал для развития, и, как описано выше, применим как для ведения здорового образа жизни, так и в качестве помощника уже заболевшим людям.

Второй сценарий заключается в предоставлении рекомендаций врачам. Врачи через рекомендации могут получить ссылки на похожие случаи заболевания, описания сопутствующих заболеваний и симптомов, ссылки на протоколы лечения и литературу, информацию о том, какие лекарственные средства дают наибольший эффект при лечении конкретного заболевания.

Приложения, следящие за здоровьем и образом жизни пользователей, могут агрегировать информацию и помогать врачам набирать участников в медицинские исследования. Таким образом, информация может собираться более регулярно, так как для измерения личных данных не надо посещать больницу, а можно направлять их исследователям напрямую из приложения. Так можно собрать более обширную и разнообразную выборку.

Существует мобильное приложение, позволяющее сфотографировать меланому и следить за ее изменением в будущем. А врачи, набрав фотографии нескольких десятков тысяч опухолей, изменяющихся во времени, могут создать алгоритм, способный распознавать злокачественную опухоль на ранних стадиях ее развития [10].

Возможен и третий сценарий предоставления рекомендаций,

в котором конечными пользователями являются организации. Кажется перспективным перенести опыт построения рекомендательных систем в торговле на фармацевтический рынок. Имея информацию о закупках больниц, можно строить ежемесячные рекомендации о том, какие препараты нужно добавить в ассортимент, чтобы повысить удовлетворенность пациентов и избежать дефицита необходимых лекарств. Информация о больничных закупках находится в открытом доступе, к этим данным применимы методы коллаборативной фильтрации, описанные в разделе 2, которые могут помочь оптимизировать больничные закупки.

4. Оптимизация больничных закупок

Процесс закупок для медицинских учреждений регулируется Федеральным законом, обязывающим публиковать сведения о закупках и их стоимости на официальных сайтах [5]. Таким образом, информация о государственных закупках находится в открытом доступе, что позволяет применять алгоритмы анализа данных и машинного обучения для автоматизации составления закупки и оптимизации ассортимента лекарственных средств, которые закупает больница. В данном разделе описана система, рекомендующая добавление в закупку больницы товаров, которые часто и в больших количествах закупают другие похожие больницы.

4.1. Постановка задачи

Рассмотрим множество больниц $S = \{S_1, \dots, S_T\}$. Каждая больница имеет известные закупки и характеризуется профилем, количеством коек, площадью, количеством корпусов, количеством отделений и профилем этих отделений, также известен адрес больницы. Таким образом, можно составить признаковое описание каждой больницы $S_i = (S_i(1), \dots, S_i(t))$. Закупки больницы S_i – это вектор уникальных товаров (лекарств и изделий медицинского назначения) $A(S_i) = \{p_1^i, \dots, p_{k(S_i)}^i\}$, где $k(S_i)$ – количество товаров в закупке. Пусть u_{ij} – количество единиц закупленного товара p_j^i .

Задача состоит в анализе совместных закупок больниц множества и построении рекомендательной системы по высоко востребованным товарам, которые следовало бы добавить в закупку конкретной больницы.

4.2. Метод построения рекомендательной системы

Закупки больниц сильно зависят от характеристик больницы, таких как терапевтический профиль больницы, ее величина. Нельзя сравнивать закупки больниц разных профилей и строить рекомендации на этом сравнении. Поэтому первым шагом решения задачи является кластеризация больниц на основе их признаков. Задача состоит в построении функции кластеризации, принимающей на вход вектор описания больницы и выдающей номер кластера, к которому эту больницу следует отнести.

В разработанном методе применяется алгоритм кластеризации k -средних [1], результат работы которого легко интерпретируем и понятен экспертам. В качестве признакового описания используются доступные характеристики больницы. Также в признаковое описание добавлено среднее количество закупленных упаковок товаров за месяц. Таким образом, кластеризация учитывает возможные ограничения склада. Результатом первого шага предложенного метода является разделение всего множества больниц на c непересекающихся кластеров:

$$S = Q_1 \cup \dots \cup Q_c,$$

где $Q_i = \cup_j S_j$, $Q_j \cap Q_i = \emptyset \quad \forall i, j$.

На втором шаге строится оценка объема закупок нового продукта больницей. Для этого данные о закупках различных больниц в кластере представляются в виде матрицы размером $n \cdot m$, где n – количество больниц в рассматриваемом кластере Q , а m – количество товаров в объединенной закупке всех больниц рассматриваемого кластера, $m = |A(S_1) \cup \dots \cup A(S_n)|$,

$S_1, \dots, S_n \in Q, |\cdot|$ – мощность множества:

$$U = \begin{bmatrix} u_{11} \dots u_{1m} \\ \vdots \quad \ddots \quad \vdots \\ u_{n1} \dots u_{nm} \end{bmatrix}.$$

Не все товары будут присутствовать во всех больницах кластера, поэтому часть ячеек матрицы будет иметь пустые значения. Интерес представляет предсказание результата в пустых ячейках для оценки потенциальности закупки нового товара больницей.

Предсказывать пустые значения матрицы U можно величиной среднего значения известных элементов матрицы по строке, или по столбцу, или некоторым комбинированным методом. Но на практике матрица U имеет большую размерность и сильно разрежена. Чтобы уменьшить влияние разреженности данных и повысить производительность метода, предлагается использовать сингулярное разложение матрицы закупок.

Методы матричного разложения позволяют найти приближение матрицы, используя матрицы меньшего ранга: матрица U приближается произведением двух матриц меньшего ранга $U \approx XY^T$, где $X \in R^{n \times k}$ – матрица признаков больниц, $Y \in R^{k \times m}$ – матрица признаков товаров, k – количество признаков, которое определяется заранее. Если размерности X и Y измеряются десятками тысяч, то k порядка нескольких десятков, таким образом, резко сокращается число параметров с nm до $(n + m)k$.

Элемент u_{ij} матрицы U приближается выражением

$$\tilde{u}_{ij} = x_i^T y_j,$$

где $x_i \in R^{k \times 1}$ – вектор признаков больницы, i -я строка матрицы X , $y_j \in R^{k \times 1}$ – вектор признаков товара, j -я строка матрицы Y .

Разложение матрицы находится путем решения задачи минимизации:

$$(1) (X^*, Y^*) = \arg \min_{X, Y} \sum_{(i,j) \in W} \left((u_{ij} - \tilde{u}_{ij})^2 + \lambda x_i^T x_i + \lambda y_j^T y_j \right),$$

где λ – параметр регуляризации. Задача оптимизации (1) решается методом градиентного спуска.

На заключительном шаге метода строятся рекомендации: больнице S_i рекомендуется добавить в закупку товары, которые ранее не закупались больницей, но имеют максимальные спрогнозированные значения \tilde{u}_{ij} .

5. Выводы

В настоящее время происходит цифровая революция в области здравоохранения: собираются электронные данные о пациентах, создаются базы специалистов-врачей, растет спрос на устройства и приложения, помогающие поддерживать здоровый образ жизни, следить за назначенным лечением, поддерживать связь с лечащим врачом.

Научный интерес представляет постановка новых задач управления и машинного обучения в области медицины. В связи с тем, что популярность и эффективность рекомендательных систем стабильно растет, повышается и актуальность адаптации математических методов разработки рекомендательных систем к новым областям, в частности разработка новых методов, решающих специфические задачи из области медицины.

В статье описан подход к разработке рекомендательной системы для оптимизации закупок лекарственных средств больницами на основе открытых данных о тендерах и закупках. Дальнейшими шагами исследования является апробация метода на реальных данных и оценка эффективности подхода.

Литература

1. ТУ ДЖ., ГОНСАЛЕС Р. *Принципы распознавания образов.* – М.: Мир, 1978.
2. <https://www.apple.com/apple-watch-series-4/health/> (дата обращения 18.10.2018).
3. <https://www.apple.com/ios/health/> (дата обращения 18.10.2018).

4. <https://today.duke.edu/2015/10/autismbeyond> (дата обращения 19.10.2018).
5. <http://zakupki.gov.ru> (дата обращения: 03.04.2018).
6. IRVING M. *One Drop: The data-driven approach to managing diabetes*. – New Atlas. – 2017.
7. JOHNS HOPKINS MEDICINE MARKETING AND COMMUNICATIONS *EpiWatch App Records Seizure Data Using Apple Watch* // Insight Tapping Innovative Solutions and Technology as Johns Hopkins Medicine. – 2015.
8. KOREN Y. *The bellkor solution to the netflix grand prize* // Netflix prize documentation.– 2009. – Vol. 81. – P. 1–10.
9. POWERS D.M. *Evaluation: from precision, recall and f-measure to roc, informedness, markedness and correlation* // J. of Machine Learning Technologies. – 2011. – Vol. 2, Iss. 1. – P. 37–63.
10. WEBSTER D.E., SUVER C., DOERR M., MOUNTS E., DOMENICO L., PETRIE T., LEACHMAN S.A., TRISTER A.D., BOT B.M. *The Mole Mapper Study, mobile phone skin imaging and melanoma risk data collected using ResearchKit* // Scientific Data. – 2017. – Vol. 4. – P. 170005.
11. WIESNER M., PFEIFER D. *Health recommender systems: concepts, requirements, technical basics and challenges* // Int. J. of Environmental Research and Public Health. – 2014. – Vol. 11(3). – P. 2580–2607.
12. XIAOYUAN SU, TAGHI M. KHOSHGOFTAAR *A survey of collaborative filtering techniques* // Advances in Artificial Intelligence archive. – 2009. – Vol. 2009. – Article No. 4.

RECOMMENDER SYSTEMS IN THE HEALTHCARE

Varvara Tsurko, V.A. Trapeznikov Institute of Control Sciences of RAS, Moscow, Cand.Sc., Researcher (v.tsurko@gmail.com).

Abstract: Recommender systems are wide spread in Internet services: contextual advertising, movies and music interesting for an user, selection of goods and services in online stores. Algorithms take into account users' profiles, history of requests and sales and the data about similar users. These algorithms develop very fast and show its effectiveness, its implementation increase the service attendance and sales. Healthcare is the new area of recommender systems application. It is collected a huge amount of medical data such as medical records, mortality and morbidity databases, treatment protocols, and even health data continuously monitored by special sensors. Ones develop intelligent systems for primary care, choosing the treatment, daily human health tracking. The article provides the general introduction to the recommender systems, the literature review of recommender systems usage in the healthcare, and also addresses the problem of building recommender systems for optimizing hospital procurement.

Keywords: recommender systems, healthcare, hospitals purchase optimization.

УДК 519.2

ББК 22.172

DOI: 10.25728/ubs.2019.82.4

*Статья представлена к публикации
членом редакционной коллегии А.И. Михальским.*

Поступила в редакцию 18.02.2019.

Дата опубликования 30.11.2019.