

М.С. АВETИСЯН,

исполнительный директор Лаборатории по искусственному интеллекту ПАО Сбербанк,
г. Москва, Россия, e-mail: avetisyan.m.s@sberbank.ru

К.С. ЕГОРОВ,

главный специалист Лаборатории по искусственному интеллекту ПАО Сбербанк,
г. Москва, Россия, e-mail: egorov.k.ser@sberbank.ru

В.Н. КОХ,

руководитель направления Лаборатории по искусственному интеллекту ПАО Сбербанк,
г. Москва, Россия, e-mail: kokh.v.n@sberbank.ru

А.А. КОШКАРОВ,

начальник информационно-вычислительного отдела, ГБУЗ «Клинический онкологический диспансер № 1» Министерства здравоохранения Краснодарского края, г. Краснодар, Россия,
e-mail: koshkarov17@yandex.ru, koshkarov@kkod.ru

Р.А. МУРАШКО,

к.м.н., главный внештатный специалист онколог Министерства здравоохранения Краснодарского края, главный врач ГБУЗ «Клинический онкологический диспансер № 1» Министерства здравоохранения Краснодарского края, г. Краснодар, Россия, e-mail: ramurashko@rambler.ru, kkod@kkod.ru

К.В. СОБЧЕНКО,

аспирант кафедры прикладной математики, ФГБОУ ВО «Кубанский государственный университет», ведущий инженер-программист информационно-вычислительного отдела ГБУЗ «Клинический онкологический диспансер № 1» Министерства здравоохранения Краснодарского края, г. Краснодар, Россия, e-mail: kostya25.06@mail.ru, k.v.sobchenko@it.kkod.ru

С.В. ШАРОВ,

к.м.н., заместитель главного врача по амбулаторно-поликлиническому разделу работы, ГБУЗ «Клинический онкологический диспансер № 1» Министерства здравоохранения Краснодарского края, г. Краснодар, Россия, e-mail: dr_sch@mail.ru, kkod@kkod.ru

А.А. ХАЛАФЯН,

д.т.н., профессор кафедры прикладной математики, ФГБОУ ВО «Кубанский государственный университет», г. Краснодар, Россия, e-mail: statlab@kubsu.ru

РАЗРАБОТКА АЛГОРИТМА ПОИСКА КЛИНИЧЕСКИ ОДНОРОДНЫХ ПАЦИЕНТОВ ПО СЛАБОСТРУКТУРИРОВАННЫМ ТЕКСТОВЫМ ДАННЫМ ЭЛЕКТРОННОЙ МЕДИЦИНСКОЙ КАРТЫ ОНКОЛОГИЧЕСКОГО ПРОФИЛЯ

УДК 519.254

Аветисян М.С., Егоров К.С., Кох В.Н., Кошкарров А.А., Мурашко Р.А., Собченко К.В., Шаров С.В., Халафян А.А. Разработка алгоритма поиска клинически однородных пациентов по слабоструктурированным текстовым данным электронной медицинской карты онкологического профиля (ПАО Сбербанк, г. Москва, Россия; ГБУЗ «Клинический онкологический диспансер № 1» министерства здравоохранения Краснодарского края, г. Краснодар, Россия; ФГБОУ ВО «Кубанский государственный университет», г. Краснодар, Россия)

Аннотация. Рост контингента пациентов со злокачественными новообразованиями в целом по России значительно увеличивает нагрузку на специализированную сеть онкологических учреждений и врачей онкологов. Наиболее вероятно, что данная тенденция сохранится в ближайшие годы. Одним из направлений повышения эффективности врачебной деятельности является извлечение современными методами анализа данных знаний из больших массивов медицинских данных путем кластеризации больных, заключающейся в выделении групп однородных (схожих) больных по совокупности клинических показателей. Целью исследования является разработка алгоритма поиска клинически однородных пациентов по слабо-

© М.С. Аветисян, К.С. Егоров, В.Н. Кох, А.А. Кошкарров, Р.А. Мурашко, К.В. Собченко, С.В. Шаров, А.А. Халафян, 2019 г.



структурированным и неструктурированным данным электронной медицинской карты онкологического диспансера, с последующей возможностью встраивания его в системы поддержки принятия врачебных решений (СППВР). Применение таких СППВР в практической медицине и в сфере медицинского образования позволит анализировать массивы информации неограниченного размера, что потребует дальнейшего внедрения и совершенствования информационных систем на всех уровнях оказания медицинской помощи. Однородность больных в работе определяли методом машинного обучения посредством косинусного расстояния в пространстве векторных представлений электронных медицинских карт. Эксперимент на 20 случайно выбранных электронных медицинских картах больных ГБУЗ «Клинический онкологический диспансер № 1» Министерства здравоохранения Краснодарского края показал высокую эффективность алгоритма в создании кластеров клинически схожих пациентов.

Ключевые слова: векторное представление, электронная медицинская карта, онкология, системы поддержки принятия врачебных решений

UDC 519.254

Avetisyan M.S., Egorov K.S., Kokh V.N., Koshkarov A.A., Murashko R.A., Sobchenko K.V., Sharov S.V., Khalafyan A.A. Development of algorithm for searching of clinically homogeneous patients from semistructured text data of oncological electronic health record (Sberbank, Moscow, Russia; Krasnodar Regional Oncological Dispenser, Krasnodar, Russia; Kuban State University, Krasnodar, Russia)

Abstract. The growth in the number of patients with malignant neoplasms in Russia significantly increases the load on a specialized network of oncological institutions and oncologists. It is most likely that this trend will continue in the coming years. One of the ways to improve the efficiency of medical activity is the extraction knowledge from medical data arrays, using modern data analysis methods, by clustering patients into groups of clinically homogeneous (similar) patients from electronic health records. The aim of the study is to develop an algorithm for finding clinically homogeneous patients according to the electronic health records of the oncological dispensary, with follow-up possibility of integration into the clinical decision support system (CDSS). The use of such CDSS in practical medicine and in the field of medical education will allow us to analyze both semistructured and unstructured arrays of information, which will require further implementation and improvement of information systems at all levels of medical care. The homogeneity of patients was determined by machine learning by cosine distance in the space of vector representations of electronic health records. An experiment on 20 randomly selected electronic health records of patients of Krasnodar Regional Oncological Dispensary showed high efficiency of the algorithm in creating clusters of clinically homogeneous patients.

Keywords: vector representation, electronic health record, oncology, clinical decision support system.

ВВЕДЕНИЕ

В своем послании Федеральному собранию 1 марта 2018 г. Президент России Владимир Владимирович Путин подчеркнул, что в России необходимо реализовать специальную общенациональную программу по борьбе с онкологическими заболеваниями, активно привлечь к решению этой задачи науку, отечественную фарминдустрию, провести модернизацию онкоцентров, выстроить современную комплексную систему, от ранней диагностики до своевременного эффективного лечения, которая позволит защитить российских граждан от тяжелых форм онкологических заболеваний, снизить смертность от рака и улучшить качество жизни пациентов, живущих со злокачественными новообразованиями (ЗНО).

В настоящее время, в РФ больны раком более 3,5 млн. человек (2,4% населения), ежегодно от онкологических заболеваний в стране умирает почти 300 тыс. человек. Онкологическая заболеваемость (интенсивные показатели) в Краснодарском крае за последние 5 лет выросла на 8,5%. В 2018 г. в крае зарегистрировано 26784 новых случаев ЗНО, что на 2803 случая больше, чем в 2014 г. На 100 тыс.

населения в 2018 г. зарегистрировано 479,4 впервые выявленных случаев ЗНО (в 2014 г. – 441,7). В РФ в 2018 г. показатель онкологической заболеваемости составил 425,4 на 100 тыс. населения страны. По уровню показателя заболеваемости ЗНО в 2017 г. Краснодарский край занимал 27-е место среди регионов РФ [3].

Актуальность темы исследования обусловлена ростом контингента пациентов с ЗНО в целом по РФ, который за последние 23 года составил 49% (1,79 млн. человек – 1993 год; 3,52 млн. человек – 2016 год), что значительно увеличивает нагрузку на специализированную сеть онкологических учреждений и врачей-онкологов. Данная тенденция сохранится в ближайшие годы и потребует дальнейшего внедрения и совершенствования информационных систем, методов машинного обучения [1], [2], [5] на всех уровнях оказания медицинской помощи.

В то же время накопленные клиниками исторические данные по результатам лечения большого количества онкобольных содержат скрытые знания в виде закономерностей по результатам лечения больных, которые можно выявить современными технологиями анализа данных. Одним из направлений добычи таких знаний является кластеризация





методами машинного обучения больных, состоящая в выделении групп клинически однородных пациентов по слабоструктурированным или неструктурированным текстовым данным электронной медицинской карты (ЭМК). Анализ таких групп больных позволит выбрать эффективную тактику лечения за счет дифференциации больных по группам однородности относительно клинического течения заболевания и учета исторического опыта их лечения. Целью исследования является разработка алгоритма поиска клинически однородных пациентов по слабоструктурированным и неструктурированным данным ЭМК онкологического диспансера, с последующей возможностью встраивания его в системы поддержки принятия врачебных решений (СППВР). Необходимость в использовании машинного обучения продиктована тем обстоятельством, что большинство методов текстового поиска не могут адекватно сравнить ЭМК клинически схожих пациентов, если они написаны разными врачами в разных медицинских учреждениях. При построении алгоритма использовали массивы данных, содержащие анонимизированные ЭМК ГБУЗ «Клинический онкологический диспансер № 1» Министерства здравоохранения Краснодарского края (КОД № 1). В рамках эксперимента для 20 случайно выбранных ЭМК разработанным алгоритмом найдено 200 наиболее схожих больных. Однородность больных определяли посредством косинусного расстояния в пространстве векторных представлений ЭМК.

МАТЕРИАЛЫ

На сегодняшний день ГБУЗ КОД № 1 является современным высокотехнологичным медицинским учреждением. Использование медицинской информационной системы (МИС) обеспечивает сохранность данных и позволяет вести статистику и анализировать структурированные массивы информации. При этом в МИС пока не разработаны инструменты для анализа неструктурированной или слабоструктурированной информации. В базе данных МИС ГБУЗ КОД № 1 с 2004 г. ведут учет всех созданных приемов и осмотров. Каждый рабочий день в ГБУЗ КОД № 1 оказывается в среднем 1300 амбулаторных поликлинических осмотров [4], ежемесячно – 3000 госпитализаций в стационарные отделения.

Для анализа и векторного представления использовали массивы данных за 14 лет эксплуатации МИС с 2004 по 2018 гг. включительно, содержащие анонимизированную информацию о поликлинических и стационарных случаях обслуживания

пациентов, о назначенных и проведенных исследованиях – 2140 518 амбулаторных осмотров; 430 558 историй болезни; 1967 806 результатов исследований.

Сформированы данные о возрасте и гендерной принадлежности пациентов, а также идентификатор сопоставления пациентов случаям обслуживания и исследованиям. Записи об исследованиях содержат такие поля как: исследование, дата результата, результат. Для амбулаторных случаев сформированы поля: дата приема, рекомендации, жалобы, анамнез заболевания, анамнез жизни, аллергологический анамнез, объективный статус, рекомендации по лечению, обследования и операции, выполненные обследования, диагноз, Международная классификация болезней (МКБ). Для стационарных случаев использовали поля: дата поступления, дата выписки, анамнез жизни, анамнез аллергологический, анамнез болезни, диагноз, МКБ, решение врачебной комиссии, лечение, причина выписки, рекомендации при выписке, модель лечения, лечебный эффект, тип лечения, статус радикального лечения, метод лечения, дата смерти, если больной умер, дата последнего осмотра, если больной жив.

Количество пациентов мужчин на 2018 г. включительно составляет 187 159 человек, количество пациентов женщин 445 604 человек. В процентном соотношении гендерное различие составляет 29,6% и 70,4% соответственно.

В десятку наиболее часто назначаемых исследований в порядке убывания по количеству назначений входят: патогистологическое исследование (ПГИ) операционного материала, общий анализ крови, общий анализ мочи, биохимическое исследование крови, электрокардиография, коагулограмма, цитологическое исследование, определение группы крови, ПГИ биопсионного материала, определение HbSAg.

Средняя длительность истории болезни стационара составляет около восьми дней. На одного пациента в среднем приходится по 4–5 поликлинических осмотров, но если задать условия выборки по пациентам, пришедшим более одного раза, то это значение возрастет до 6–7.

По данным МИС 73,4% диагнозов, выставленных в ГБУЗ КОД № 1, относятся к группе «С» по МКБ 10-го пересмотра (МКБ-10). Также значительную часть от числа выставляемых диагнозов составляют группы «N» и «D», по 12,3% и 12,1% соответственно. Далее следуют более малочисленные группы, составляющие менее 1% от общего числа каждый: «E», «Z», «K», «J», «O», «I», «L» и другие.



МЕТОДЫ

Описание случаев заболевания в электронной медицинской карте (ЭМК) пациента ГБУЗ КОД № 1 представляет собой, в основном, неструктурированные или слабоструктурированные записи врачей на естественном языке с обилием медицинских терминов и сокращений. Большинство методов текстового поиска без применения машинного обучения не могут адекватно сравнить ЭМК клинически однородных пациентов, если они написаны разными врачами в разных медицинских учреждениях. Для достижения сформулированной ранее цели методом обработки естественного языка (NLP, natural language processing) построены векторные представления ЭМК. Полученные вектора сравнивались между собой посредством косинусного расстояния, которое в области NLP работает лучше эвклидова из-за того, что в начале обучения векторные представления слов инициализируются небольшой случайной величиной, и вектор растет в процессе обучения. Это приводит к тому, что слова, встречающиеся чаще, могут иметь большую длину вектора, чем слова встречающиеся реже, что вносит ошибку в определении семантической близости слов. Векторное представление ЭМК представляет собой сумму векторных представлений слов, входящих в него, что тоже может внести ошибку при сравнении ЭМК с разным количеством слов. Косинусное расстояние учитывает только направление вектора, что устраняет необходимость в нормализации векторных представлений. Кроме того, большим преимуществом данного метода является его универсальность – имея хорошие векторные

представления ЭМК можно эффективно решать и другие прикладные задачи – классификацию, кластеризацию пациентов, подбор лечения, предсказание исходов лечения и другие.

Векторные представления ЭМК были получены несколькими способами:

1. Использовали предобученные векторы слов, полученные с помощью модели Continuous Bag Of Words (CBOW) [10] на текстах CommonCrawl.org и Wikipedia.org [8].

2. Используя библиотеку FastText [9] выучены векторные представления без учителя по схеме Continuous Skip-gram Model (рис. 1) на данных ГБУЗ КОД № 1.

При такой схеме модель учится предсказывать контекст по слову. Например, если взять предложение “кожный покров и слизистые физиологической окраски”, то модель будет учиться увеличивать вероятность слов “кожный”, “покров”, “физиологической” и “окраски”, если на вход подать слово “слизистые”.

Для обучения берется участок текста длиной в 5 слов. На вход модели подается центральное (3-е) слово участка, затем слово преобразуется в вектор размерности D из таблицы $V \times D$, где V – размер словаря. Кроме матрицы $V \times D$ одновременно обучается матрица $D \times V$ таким образом, что при матричном произведении вектора слова D на матрицу $D \times V$ получается вектор V , в котором значения, соответствующие индексам слов w_1, w_2, w_4 и w_5 максимальны, а остальные – минимальны. Такой вектор проходит через функцию (1), получая вероятности слов контекста, затем вычисленные вероятности сравниваются с истинным распределением слов

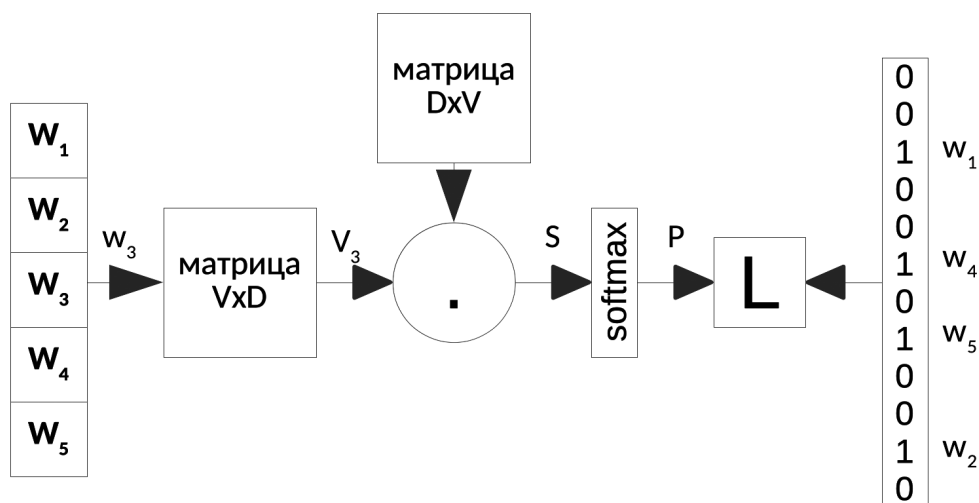


Рис. 1. Модель Continuous Skip-gram.



контекста (где вероятности слов w_1, w_2, w_4, w_5 равны 1, а остальные равны 0). В качестве функции оценки ошибки L использовали логистическую функцию потерь (2).

$$\text{softmax}(x)_i = \frac{\exp(x_i)}{\sum_j \exp(x_j)} \quad (1)$$

$$L(y, p) = \sum_{i=1}^N \sum_{j=1}^C y_{ij} \ln(p_{ij}) \quad (2)$$

3. Третий подход – используя ту же библиотеку FastText, но уже метод обучения с учителем [6], где в качестве цели классификации определены группы заболеваний по нозологиям, т.е. модель училась предсказывать диагноз с точностью до группы заболеваний по ЭМК. Схема обучения с учителем представлена на рис. 2.

Слова ЭМК подаются на вход модели, затем для каждого слова из матрицы $V \times D$ берется его нормализованное векторное представление и все они суммируются. Полученный вектор умножается на матрицу $D \times C$ и получается вектор размерности C (где C – количество классов), в котором максимальное значение должно соответствовать правильному классу (диагнозу). Этот вектор затем пропускается через функцию и сравнивается с истинным вектором класса (где компонента истинного класса равна 1, а все остальные компоненты равны 0). В качестве функции оценки ошибки L аналогично использовали логистическую функцию потерь.

Стоит заметить, что приведенное выше описание метода соответствует классическому подходу. Учитывая специфику данных и решаемой задачи, при построении алгоритма нами были внесены незначительные изменения. В частности, при обучении без учителя, вместо классической функции,

был использован иерархический [7]. Кроме того, векторное представление слова состояло из усредненных векторов n -грамм слова [9]. Например, если взять n -граммы слова величины 3, то векторное представление слова “диагноз” будет равно среднему от векторных представлений самого слова и всех n -грамм этого слова: <ди, диа, иаг, агн, гно, ноз, оз>, где символы < и > – служебные символы, означающие конец и начало слова. В этом случае полученные вектора слов более устойчивы к опечаткам и сокращениям, можно получить адекватное векторное представление слова, не встречавшееся в обучающей выборке. Также полученные вектора могут учитывать морфологию слова, т.к. модель определяет приставки и окончания слов. В настоящей работе использовали n -граммы слов размера 3, 4 и 5.

Как альтернативу методу обработки естественного языка, были применены рекуррентные нейронные сети, но они показали более низкие результаты, возможно из-за недостаточного объема обучающей выборки.

РЕЗУЛЬТАТЫ И ОБСУЖДЕНИЕ

При сравнении качества полученных представлений исходили из гипотезы, что к однородным пациентам применялось похожее лечение. Кроме векторных представлений ЭМК, построены векторные представления описания лечения пациентов. Таким образом, возможно сравнение созданных моделей поиска клинически однородных пациентов по следующему алгоритму:

1. Выбираем случайного пациента.
2. Определяем косинусное расстояние между ЭМК выбранного пациента и остальными пациентами.

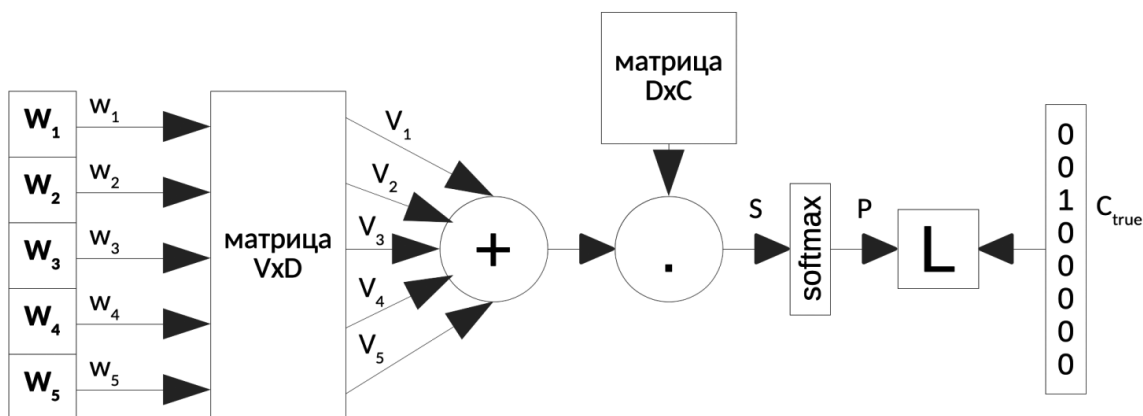


Рис. 2. Обучение с учителем.



3. Определяем косинусное расстояние между описанием назначенного лечения выбранного пациента и остальными пациентами.

4. Отбираем 20 пациентов, наиболее близких выбранному по ЭМК.

5. Получаем оценку близости отобранных пациентов к исходному усредняя их расстояния по описанию лечения до исходного пациента.

6. Повторяем пп. 1–5 с 50 случайными пациентами.

7. Усредняем оценки близости 50-ти пациентов.

$$S = \frac{1}{NK} \sum_{i=1}^N \sum_{j=1}^K D_{ij} \quad (3)$$

Полученная средняя оценка близости (3), где $N=50$ – количество тестовых пациентов, $K=20$ – количество ближайших пациентов к данному тестовому, D_{ij} – косинусное расстояние между векторными представлениями описания лечения i -го тестового пациента и j -го наиболее близкого к нему исходя из расстояния между векторными представлениями ЭМК, является критерием сравнения двух алгоритмов поиска клинически однородных пациентов.

Изначально была выбрана модель поиска векторных представлений описаний лечения.

В *таблице 1* представлены средние оценки близости для модели с предобучением и без предобучения. Из результатов видно, что предобученная модель на порядок превосходит модель без предобучения.

Затем с помощью лучшей модели поиска векторных представлений описаний лечения по вышеприведенному алгоритму были протестированы различные модели поиска векторных представлений ЭМК. Результаты представлены в *таблице 2*.

Видна тенденция в увеличении качества поиска однородных пациентов с увеличением размерности векторного представления, кроме того неоспоримо повышение качества при использовании предобученных моделей.

По предложенному критерию наилучшую из моделей протестировали в условиях врачебной практики (обучения). Разработанный алгоритм поиска клинически однородных пациентов показал свою жизнеспособность на примере случайно выбранных 20 пациентов. Найденные алгоритмом по 10 клинически наиболее однородных пациентов к каждому из них отображены в *таблице 3*.

Иллюстрация реальных примеров поиска клинически однородных пациентов в МИС ГБУЗ КОД № 1 представлена на *рис. 3, 4*.

Таблица 1

Сравнение моделей поиска векторных представлений описания лечения

Размерность векторного представления	Предобучение на открытых данных	Оценка близости
300	нет	0.156
300	да	0.012

Таблица 2

Сравнение моделей поиска векторных представлений ЭМК

С учителем / без учителя	Предобучение на открытых данных	Размерность векторного представления	Оценка близости
б/у	да	300	0.0196
б/у	нет	25	0.0207
б/у	нет	50	0.0195
б/у	нет	100	0.0193
б/у	нет	200	0.0197
б/у	нет	300	0.0194
с/у	да	300	0.0188
с/у	нет	25	0.0200
с/у	нет	50	0.0200
с/у	нет	100	0.0199
с/у	нет	200	0.0201
с/у	нет	300	0.0196



Таблица 3

**Идентификаторы выбранных ЭМК для эксперимента
и результат поиска наиболее однородных**

ID ЭМК	1	2	3	4	5	6	7	8	9	10	
	95031	774790	815771	872109	886047	932660	977866	998040	998524	1000210	
Наиболее похожие ЭМК	1	586925	780424	86694	970286	816580	864332	969185	986890	984047	682577
	2	877824	754608	744909	907965	514454	655422	973826	982209	94822	710333
	3	636047	554007	549734	736317	859086	556490	587227	589959	948373	952068
	4	630794	611605	828958	791429	32772	812840	969673	942571	650492	955842
	5	966368	98135	810236	909578	900821	909434	922590	629956	481956	900864
	6	627075	753936	783482	644313	941701	931255	960616	604108	475477	904014
	7	122374	568506	478589	647716	920686	509734	933159	557740	769319	583161
	8	26118	742601	878487	958733	609722	642455	657909	560958	600463	690966
	9	541446	103949	833489	747371	928180	603373	958565	656950	480477	963054
	10	939064	595320	72624	38305	841731	568976	669464	982994	522534	699018
ID ЭМК	11	12	13	14	15	16	17	18	19	20	
	1001235	1001359	1001639	1001742	1002143	1003838	1004218	1004480	1004838	1005167	
Наиболее похожие ЭМК	1	990659	840188	816111	680621	984789	678431	925650	985720	724194	679462
	2	982204	773311	702463	763376	975936	902092	956637	978611	748672	636827
	3	589463	725014	865242	486380	933724	630403	958408	978718	605681	696646
	4	978949	841143	707922	925463	989961	719372	965284	895203	655058	728896
	5	703331	936057	568665	115609	952127	907520	927730	989075	721327	725831
	6	114774	865455	732485	511599	973353	664981	762446	981325	726178	882661
	7	973078	713023	669399	738839	924503	736856	565975	661586	484120	741698
	8	770505	752928	624043	886150	968945	755769	982443	618059	987501	641507
	9	968389	663021	888197	838393	965291	685983	961118	986098	712444	623237
	10	980407	687069	563908	86592	990763	667732	942323	982425	747553	978746

Осмотр пациента

Общее Предпосылки Текущее состояние Рекомендации Результаты

Состояние (шаблоны)

По шаблону
Общее состояние удовлетворительное. Кожные покровы нормального цвета, геморагий нет. Дыхание везикулярное. Сердечные сокращения приглушены, ритмичные. Живот не увеличен.

Давление Температура Пульс ЧДД

Шкала Карновского 100 % ОКС ECG: 0

Status localis (шаблоны)
периф лу не ув, гепатоспленомегалии нет

Установленный диагноз Сопутствующие заболевания и осложнения Метод подтверждения диагноза

C90 Множественная миелома и злокачественные плазмоклеточные новообразования

С90.0 Множественная миелома

Для печати Множественная миелома, с секретией Ig G, 3 А ст.

TNM Стадия

Доп. информация после 15 курсов ПХТ (7 VCD , 5 CP , 4 MP), сост Клиническая группа II

сохранить данный документ как шаблон

Сохранить Назад Далее Отменить

Рис. 3. Эталонная ЭМК пациента.



Рис. 4. Пример ЭМК наиболее клинически однородного пациента, найденная посредством разработанного алгоритма.

На рисунках изображены вкладки текущего состояния в окне осмотра пациента. Диагнозы по МКБ-10, клиническая группа совпадают (лечащие врачи разные). У найденного клинически однородного пациента установлена стадия.

ВЫВОДЫ

В результате исследования на основе векторного представления ЭМК разработан алгоритм поиска клинически однородных пациентов по совокупности клинических показателей течения заболевания.

Наилучшая из моделей векторного представления ЭМК протестирована в условиях врачебной практики. Эксперимент на 20 случайно выбранных ЭМК больных ГБУЗ КОД № 1 показал высокую эффективность алгоритма в создании кластеров клинически однородных пациентов.

Анализ групп схожих пациентов позволит выбрать эффективную тактику лечения за счет дифференциации больных по группам однородности и учета исторического опыта их лечения.

Совокупность полученных результатов говорит о целесообразности экспериментального внедрения разработанного алгоритма в МИС ГБУЗ КОД № 1 и другие клиники онкологического профиля для поддержки принятия врачебных решений: при

постановке диагноза, выборе тактики лечения, формирования автоматизированных шаблонов описаний на основе предыдущих случаев лечения, хранящихся в базе данных.

В образовательных целях разработанный алгоритм может быть реализован как автономное приложение для помощи студентам и слушателям курсов повышения квалификации при симуляции ситуаций в медицине на реальных анонимизированных исторических данных онкологических больных.

Развивать разработанный алгоритм имеет смысл, в первую очередь, в сторону усложнения моделей. При увеличении масштаба сотрудничества медицинских учреждений и лабораторий искусственного интеллекта, возможно, за счет значительного увеличения объемов медицинских данных более перспективным методом решения задачи кластеризации пациентов станут искусственные нейронные сети.

Очевидно, предложенный подход к построению алгоритмов поиска клинически однородных пациентов по неструктурированным и слабо структурированным данным электронной медицинской карты может быть применен в различных областях здравоохранения.



ЛИТЕРАТУРА



1. Гусев А.В. Перспективы нейронных сетей и глубокого машинного обучения в создании решений для здравоохранения // Врач и информационные технологии. – 2017. – № 3. С. 92–105.
2. Кузнецова А.В., Сенько О.В., Кузнецова Ю.О. Преодоление проблемы «черного ящика» при использовании методов машинного обучения в медицине // Врач и информационные технологии. – 2018. – Специальный выпуск Труды Международного конгресса «Информационные технологии в медицине 2018» 11–12 октября, 2018 г. С. 74–80.
3. Мурашко Р.А., Тесленко Л.Г., Цокур И.В., Степанова Л.Л., Белокрылова А.П., Попкова А.Г., Майковская А.Г. Злокачественные новообразования в Краснодарском крае. Состояние онкологической помощи населению // Под ред. доцента кафедры онкологии с курсом торакальной хирургии ФПК и ППС ФГБОУ ВО КубГМУ Минздрава России, кандидата медицинских наук Р.А. Мурашко. – Краснодар, 2019.
4. Собченко К.В., Коваленко А.В., Кошкарров А.А., Мурашко Р.А., Шаров С.В. Разработка алгоритма автоматизированного вейвлет-анализа данных о работе регистратуры клинического онкологического диспансера на региональном уровне // Врач и информационные технологии. – 2018. – Специальный выпуск Труды Международного конгресса «Информационные технологии в медицине 2018» 11–12 октября, 2018 г. С. 66–73.
5. Халафян А.А., Карахалис Л.Ю., Папова Н.С., Акиньшина В.А., Кошкарров А.А. Прогнозирование стадии распространения заболевания у пациентов, страдающих аденомиозом, нейронными сетями // Врач и информационные технологии. – 2018. – № 4. С. 64–74.
6. Joulín A., Grave E., Bojanowski P., Mikolov T. Bag of Tricks for Efficient Text Classification.
7. Frederic Morin, Yoshua Bengio. Hierarchical Probabilistic Neural Network Language Model.
8. Grave Edouard, Bojanowski Piotr, Gupta Prakhar, Joulín Armand, Mikolov Tomas. Learning Word Vectors for 157 Languages.
9. Bojanowski P., Grave E., Joulín A., Mikolov T. Enriching Word Vectors with Subword Information.
10. Mikolov Tomáš, Chen Kai, Corrado Greg D., Dean Jeffrey. Efficient estimation of word representations in vector space //arXiv preprint, 2013a arXiv:1301.3781.

Новости отрасли

КАРТА «ИСКУССТВЕННЫЙ ИНТЕЛЛЕКТ
В ЗДРАВООХРАНЕНИИ РОССИИ»

Искусственный интеллект является одним из самых обнадеживающих направлений развития цифрового здравоохранения. У нас в стране постоянно появляются все новые и новые разработки и исследования, предлагающие реализацию этих технологий для медицины и здравоохранения.

Группа исследователей из ассоциации разработчиков и пользователей систем искусственного интеллекта для медицины «Национальная база медицинских знаний» (НБМЗ) собрала информацию об известных на данный момент примерах внедрения и применения систем на базе ИИ в здравоохранении субъектов РФ. В итоге исследователи обнаружили, что в 22 регионах такие проекты уже запущены и некоторые из них развиваются самым активным образом.

Обзор и карта таких проектов доступны по адресу: <https://webiomed.ai/blog/karta-iskusstvennyi-intellekt-v-zdravookhraneni-rossii/>

