



(51) МПК
G06F 17/00 (2006.01)
G16H 10/60 (2018.01)
G16H 50/20 (2018.01)
G06N 3/02 (2006.01)
G06N 20/00 (2019.01)

ФЕДЕРАЛЬНАЯ СЛУЖБА
 ПО ИНТЕЛЛЕКТУАЛЬНОЙ СОБСТВЕННОСТИ

(12) ОПИСАНИЕ ИЗОБРЕТЕНИЯ К ПАТЕНТУ

(52) СПК

G06F 17/00 (2019.02); *G16H 10/60* (2019.02); *G16H 50/20* (2019.02); *G06N 3/02* (2019.02); *G06N 20/00* (2019.02)

(21)(22) Заявка: 2017137802, 29.12.2017

(24) Дата начала отсчета срока действия патента:
29.12.2017

Дата регистрации:
21.10.2019

Приоритет(ы):

(22) Дата подачи заявки: 29.12.2017

(43) Дата публикации заявки: 01.07.2019 Бюл. № 19

(45) Опубликовано: 21.10.2019 Бюл. № 30

Адрес для переписки:

105037, Москва, ул. Первомайская, 4, кв. 43,
Абраменко О.И.

(72) Автор(ы):

Дрокин Иван Сергеевич (RU),
Бухвалов Олег Леонидович (RU),
Сорокин Сергей Юрьевич (RU)

(73) Патентообладатель(и):

Общество с ограниченной ответственностью
"Интеллоджик" (RU)

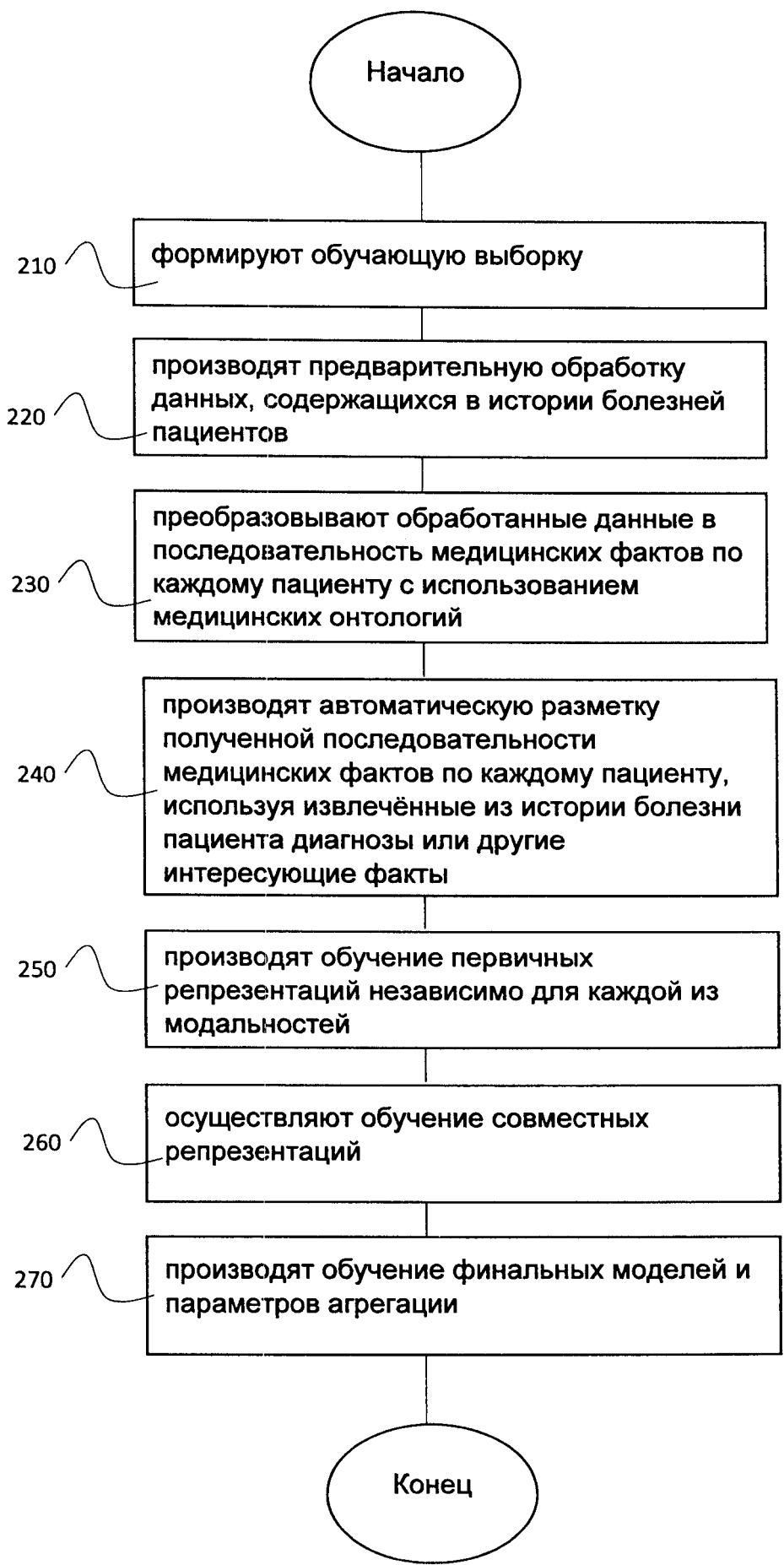
(56) Список документов, цитированных в отчете о поиске: US 7630947 B2, 08.12.2009. US 7899764 B2, 01.03.2011. US 20090083231 A1, 26.03.2009. US 8015136 B1, 06.09.2011. US 2006/020444 A1, 26.01.2006. RU 2606050 C2, 10.01.2017. RU 2491622 C1, 27.08.2013.

(54) Способ и система поддержки принятия врачебных решений с использованием математических моделей представления пациентов

(57) Реферат:

Изобретение относится к области искусственного интеллекта. Технический результат - повышение точности диагностирования, анализа и прогноза развития заболевания для пациента. Обрабатывают данные, содержащиеся в истории болезней пациентов, выбранных из предварительно сформированной обучающей выборки; преобразовывают эти данные в последовательность медицинских фактов по каждому пациенту с использованием медицинских онтологий; производят автоматическую разметку полученной последовательности медицинских фактов по каждому пациенту, используя извлеченные из истории болезни пациента диагнозы или другие интересующие факты; производят обучение первичных репрезентаций

независимо для каждой из модальностей; осуществляют обучение совместных репрезентаций; производят обучение финальных моделей и параметров агрегации; получают историю болезни пациента, не входящую в обучающую выборку, и производят предварительную обработку данных из неё; преобразовывают обработанные данные в последовательность медицинских фактов с использованием медицинских онтологий; полученный набор фактов отправляют на вход обученным финальным моделям; определяют диагноз, проводят анализ и прогноз развития заболеваний пациента с наибольшей вероятностью, соответствующий предъявленному набору фактов. 2 н. и 1 з.п. ф-лы, 11 ил.



Фиг.2



FEDERAL SERVICE
FOR INTELLECTUAL PROPERTY

(51) Int. Cl.
G06F 17/00 (2006.01)
G16H 10/60 (2018.01)
G16H 50/20 (2018.01)
G06N 3/02 (2006.01)
G06N 20/00 (2019.01)

(12) **ABSTRACT OF INVENTION**

(52) CPC

G06F 17/00 (2019.02); *G16H 10/60* (2019.02); *G16H 50/20* (2019.02); *G06N 3/02* (2019.02); *G06N 20/00* (2019.02)

(21)(22) Application: **2017137802, 29.12.2017**(24) Effective date for property rights:
29.12.2017Registration date:
21.10.2019

Priority:

(22) Date of filing: **29.12.2017**(43) Application published: **01.07.2019** Bull. № 19(45) Date of publication: **21.10.2019** Bull. № 30

Mail address:

**105037, Moskva, ul. Pervomajskaya, 4, kv. 43,
Abramenko O.I.**

(72) Inventor(s):

**Drokin Ivan Sergeevich (RU),
Bukhvalov Oleg Leonidovich (RU),
Sorokin Sergej Yurevich (RU)**

(73) Proprietor(s):

**Obshchestvo s ogranichennoj otvetstvennostyu
"Intellodzhik" (RU)**(54) **METHOD AND SYSTEM FOR SUPPORTING MEDICAL DECISION MAKING USING MATHEMATICAL MODELS OF PRESENTING PATIENTS**

(57) Abstract:

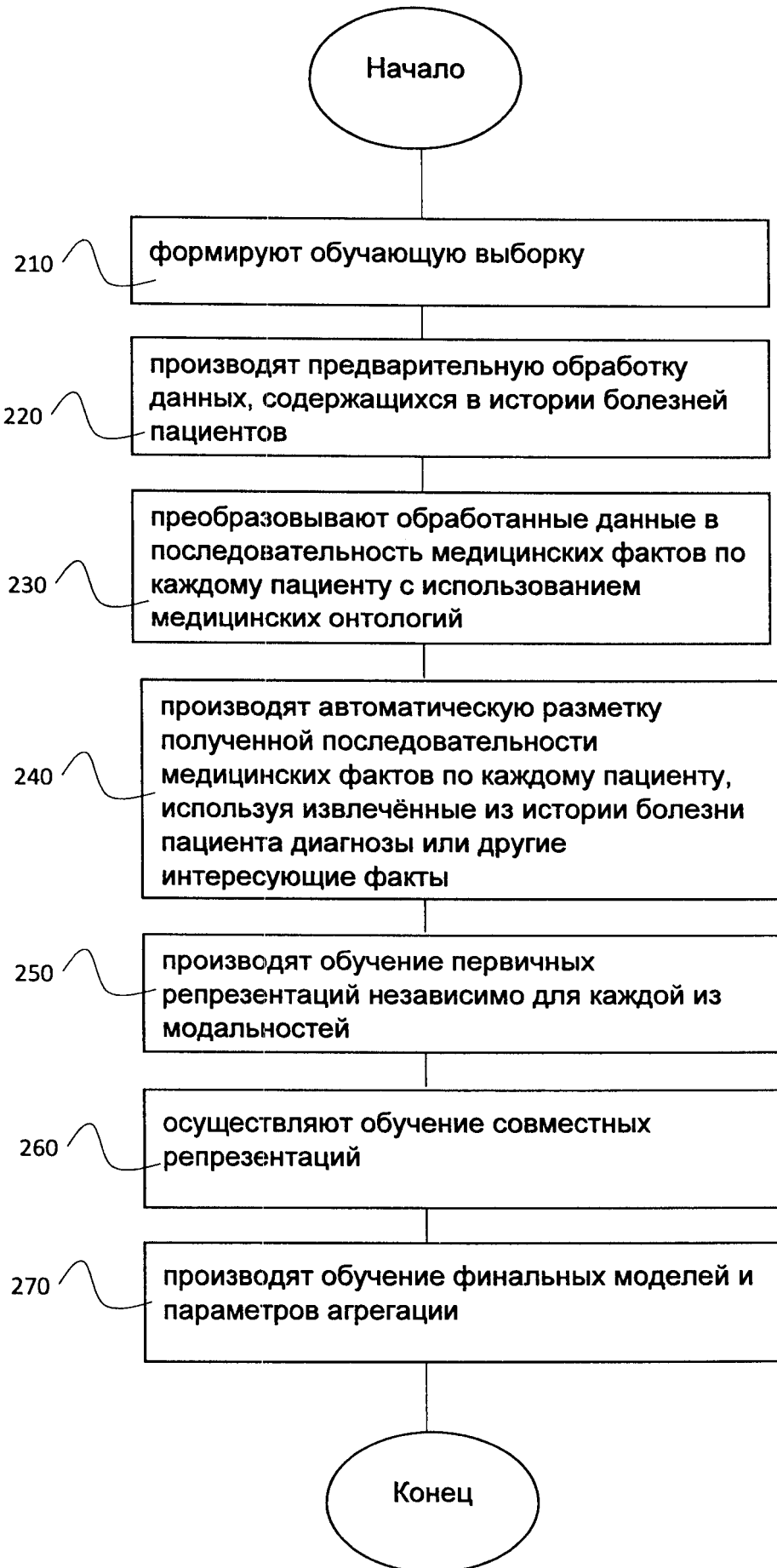
FIELD: physics.

SUBSTANCE: invention relates to the field of artificial intelligence. Processing data contained in patient's medical history selected from pre-formed training sample; converting said data into a sequence of medical facts for each patient using medical ontologies; performing the obtained sequence of medical facts automatic marking for each patient using the diagnoses or other facts of interest extracted from the patient's medical history; training primary representations independently for each of the modalities; training of joint representations; final model and aggregation parameters training; patient's medical

history not included in the training sample is obtained, and preliminary data processing is performed; converting processed data into a sequence of medical facts using medical ontologies; obtained set of facts is sent to input to trained final models; diagnosis is determined, analysis and prognosis of patient's diseases development is performed with highest probability corresponding to presented set of facts.

EFFECT: technical result is higher accuracy of diagnosis, analysis and prediction of developing disease for the patient.

3 cl, 11 dwg



Фиг.2

Область техники

Данное техническое решение относится к области искусственного интеллекта, а именно к системам поддержки принятия решений в медицине.

Уровень техники

5 Выявление и диагностирование болезней пациента является сложной задачей, которой обычно занимаются врачи. Существует множество факторов, которые могут повлиять на результат диагностирования врачом - опыт врача, внимательность, сложность текущего случая. Для устранения этих недостатков используются различные системы поддержки принятия решений.

10 В большинстве подходов система поддержки принятия решения позволяет выявить наличие или отсутствие некоторых патологий (болезней), но не могут проводить анализ и прогноз развития заболеваний для конкретного пациента

Сущность

15 Данное техническое решение позволяет создать математическую модель пациента при помощи которой становится возможно повысить точность диагностирования и осуществлять анализ и прогноз развития заболеваний для конкретного пациента.

Способ поддержки принятия врачебных решений с использованием математических моделей представления пациентов, выполняемая на сервере, включает следующие шаги: формируют обучающую выборку, содержащую электронную историю болезни

20 пациентов, сгруппированных по пациенту;

производят предварительную обработку данных, содержащихся в истории болезней пациентов, выбранных из обучающей выборки;

преобразовывают обработанные данные в последовательность медицинских фактов по каждому пациенту с использованием медицинских онтологий;

25 производят автоматическую разметку полученной последовательности медицинских фактов по каждому пациенту, используя извлеченные из истории болезни пациента диагнозы или другие интересующие факты;

производят обучение первичных репрезентаций независимо для каждой из модальностей;

30 осуществляют обучение совместных репрезентаций;

производят обучение финальных моделей и параметров агрегации;

получают историю болезни пациента, не входящую в обучающую выборку;

производят предварительную обработку данных полученной истории болезни пациента;

35 преобразовывают предварительно обработанные данные в последовательность медицинских фактов с использованием медицинских онтологий;

полученный набор фактов отправляют на вход сформированным моделям;

определяют диагноз, а также проводят анализ и прогноз развития заболеваний пациента с наибольшей вероятностью соответствующий предъявленному набору фактов.

40 Система поддержки принятия врачебных решений с использованием математических моделей представления пациентов включает, по крайней мере, один процессор, оперативную память, запоминающее устройство, содержащее инструкции, загружаемые в оперативную память и выполняемые, по крайней мере, одним процессором, инструкции содержат следующие шаги:

45 формирование обучающей выборки, содержащую электронную историю болезни пациентов, сгруппированных по пациенту;

предварительную обработку данных, содержащихся в истории болезней пациентов, выбранных из обучающей выборки;

преобразование обработанных данные в последовательность медицинских фактов по каждому пациенту с использованием медицинских онтологий;

автоматическую разметку полученной последовательности медицинских фактов по каждому пациенту, используя извлеченные из истории болезни пациента диагнозы или другие интересующие факты;

обучение первичных репрезентаций независимо для каждой из модальностей;

обучение совместных репрезентаций;

обучение финальных моделей и параметров агрегации;

получение истории болезни пациента, не входящей в обучающую выборку;

предварительную обработку данных полученной истории болезни пациента;

преобразование предварительно обработанных данных в последовательность медицинских фактов с использованием медицинских онтологий;

отправку полученного набор фактов на вход сформированным моделям;

определение диагноза, а также анализа и прогноза развития заболеваний пациента с наибольшей вероятностью соответствующий предъявленному набору фактов.

В одном из своих широких аспектов, данное техническое решение позволяет моделировать процессы и тенденции в организме пациента, выявлять влияние медикаментов и назначенного лечения, определять смертность пациента после операций или назначения лечения.

Подробное описание

Ниже перечислены термины и определения, используемые в данном техническом решении.

Векторное представление - общее название для различных подходов к моделированию языка и обучению представлений в обработке естественного языка, направленных на

сопоставление словам (и, возможно, фразам) из некоторого словаря векторов из R^n для n , значительно меньшего количества слов в словаре. Теоретической базой для векторных представлений является дистрибутивная семантика. Существует несколько методов для построения такого сопоставления, например, используют нейронные сети, методы понижения размерности в применении к матрицам совместных упоминаний слов (англ. word co-occurrence matrices) и явные представления, обучающиеся на контекстах упоминаний слов (англ. explicit representations).

Векторное представление пациента - математическая модель пациента, на основании физиологических параметров пациента, анамнеза, истории болезней и их лечения (методов и хода лечения, прописанных препараты и т.п.) и т.д., позволяющая прогнозировать развития заболеваний, диагностировать, формировать рекомендации, стратегии лечения и т.д. для конкретного пациента.

Гистология - раздел биологии, изучающий строение, жизнедеятельность и развитие тканей живых организмов.

Гистология человека - раздел медицины, изучающий строение тканей человека.

Дистрибутивная семантика - это область лингвистики, которая занимается вычислением степени семантической близости между лингвистическими единицами на основании их распределения (дистрибуции) в больших массивах лингвистических данных (текстовых корпусах). Дистрибутивная семантика основывается на дистрибутивной гипотезе: лингвистические единицы, встречающиеся в схожих контекстах, имеют близкие значения.

Метаданные - информация о другой информации, или данные, относящиеся к дополнительной информации о содержимом или объекте. Метаданные раскрывают сведения о признаках и свойствах, характеризующих какие-либо сущности, которые

позволяют автоматически искать и управлять ими в больших информационных потоках.

5 Модальность данных - принадлежность данных к некоторому источнику данных, определяющему структуру упомянутых данных, их формат, а также позволяющему соотнести упомянутую структуру с той или иной системой органов и/или нозологий и/или процедур. В качестве источника данных могут выступать как средства, получения данных о пациенте, так и сам пациент.

10 Онтология - всеобъемлющая и детальная формализация некоторой области знаний с помощью концептуальной схемы. Обычно такая схема состоит из иерархической структуры данных, содержащей все релевантные классы объектов, их связи и правила (теоремы, ограничения), принятые в этой области.

15 Регуляризация (в статистике, машинном обучении, теории обратных задач) - метод добавления некоторой дополнительной информации к условию с целью решить некорректно поставленную задачу или предотвратить переобучение. Эта информация часто имеет вид штрафа за сложность модели, например, это могут быть ограничения гладкости результирующей функции или ограничения по норме векторного пространства.

Стемминг (англ. stemming) - процесс нахождения основы слова для заданного исходного слова. Основа слова необязательно совпадает с морфологическим корнем слова.

20 Факт (медицинский факт) - данные, описывающие пациента, в том числе способы его лечения и связь упомянутых данных с другими медицинскими фактами.

25 Электронная история болезней пациентов (электронная медицинская карта, электронный паспорт пациента; англ. electronic medical record - EMR, electronic health record) - база данных, содержащая сведения о пациенте: физиологические параметры пациента, анамнез, истории болезней и их лечение (методы и ход лечения, прописанные препараты и т.п.). В том числе электронная история болезней пациентов содержит записи пациентов, включающих, по меньшей мере, следующие данные: дату добавления записи, коды диагнозов, симптомов, процедур и лекарств, текстовое описание истории болезни на естественном языке, ассоциированные с историей болезни биомедицинские изображения, результаты исследований и анализов пациентов.

30 Медицинский персонал или другой пользователь медицинской информационной системы загружает на сервер через ее интерфейс данные о пациенте, которые содержат историю болезни пациента, сведения о физиологических параметрах пациента, другую информацию.

35 Еще в одном из вариантов реализации данные на сервер могут быть загружены автоматически, без участия человека, например, при сборе и исследовании анализов, проведения лечащих процедур и т.д.

40 При обращении пациента к врачу, прохождении обследования, сдачи анализов или других медицинских процедурах, в медицинской информационной системе формируются (заполняются и сохраняются) данные по каждой такой процедуре. Данные могут включать записи осмотров пациента, коды диагнозов, симптомов, процедур и лекарств, назначаемых и/или принимаемых пациентом, описание истории болезни на естественном языке, биомедицинские изображения, результаты анализов, исследований, результаты наблюдений/измерений физиологических параметров, ЭКГ, ЭЭГ, МРТ, УЗИ, биопсия, цитологические исследования, рентгена, маммографии, но не ограничиваясь указанными данными.

45 Указанные данные могут быть представлены в текстовом, табличном формате, в виде временных рядов, изображений, видео, геномных данных, сигналов, но не

ограничиваясь. Данные также могут быть представлены в структурированном и неструктурированном виде. Дополнительно в качестве данных могут выступать связи между приведенными выше данными.

5 Анализы могут включать, но не ограничиваясь, анализ крови, спинномозговой жидкости, мочи, кала, генетические тесты и т.д. В рамках технического решения не накладывает ограничения на типы анализов.

Рассмотрим пример, проиллюстрированный на Фиг. 1.

10 Один пациент 105 приходит на первичный осмотр к профильному врачу. Врач производит необходимые медицинские действия, после чего формирует описание имеющихся у пациента симптомов, дает назначения на проведение анализов. Далее врач вводит данную информацию в компьютер через интерфейс медицинской информационной системы 110, после чего эти данные сохраняются в электронной истории болезни. Другой пациент 106 приходит на повторный прием к врачу-терапевту. Врач-терапевт производит назначение лекарственных препаратов пациенту, внося эти
15 данные в электронную медицинскую карту. Таким образом, по каждому пациенту формируется необходимый набор записей в электронных историях болезни, который в дальнейшем может использоваться другими врачами или системами поддержки принятия решений.

20 Для построения систем поддержки принятия решений в области медицины требуется набрать определенный объем данных, позволяющий обучить систему распознавать по полученным медицинским данным диагнозы, группы диагнозов или фактов. Когда необходимый объем данных собирается в медицинской информационной системе, он может быть использован в качестве обучающей выборки. Большинство существующих систем поддержки принятия решений используют в своей составляющей машинное
25 обучение (англ. machine learning) в различных его проявлениях.

Одним из важных компонентов в системе поддержки принятия медицинских решений является векторное представление пациента (математическая модель пациента), позволяющее прогнозировать развития заболеваний, диагностировать, формировать
30 рекомендации, стратегии лечения и др. для конкретного пациента.

30 В одном из вариантов реализации функционал, позволяющий формировать векторное представление пациента, располагается на отдельном сервере.

В одном из вариантов реализации функционал, позволяющий формировать векторное представление пациента, располагается на том же сервере, где и располагается
35 медицинская информационная система.

35 В одном из вариантов реализации функционал, позволяющий формировать векторное представление пациента, представляет собой облачный сервис (англ. cloud service), использующий облачные вычисления (англ. cloud computing) на распределенной системе серверов.

40 На первом этапе для формирования векторного представления пациента формируют обучающую выборку 210, фиг. 2 (англ. training dataset), которая в дальнейшем будет использована для обучения алгоритмами машинного обучения, в том числе алгоритмами глубокого обучения (англ. deep learning).

В одном из вариантов реализации обучающая выборка формируется пользователем
45 медицинской информационной системы, путем отбора записей пациентов.

45 В одном из вариантов реализации отбор записей может осуществляться по заданным критериям. При этом в качестве упомянутых критериев могут выступать по меньшей мере:

- правила включения/выключения в обучающую выборку:

○ пациентов с анамнезом из заданной совокупности анамнезов (Например, только пациенты с онкологическим анамнезом);

○ пациентов, удовлетворяющие заданным тендерным или возрастным параметрами (Например, только мужчины в возрасте от 30 до 45 лет);

5 ○ пациентов, связанных с пациентами, уже включенными в обучающую выборку, при этом связь определяется по меньшей мере схожестью анамнезов, способов лечения и т.д.

- правила включения сформированных ранее обучающих выборок.

10 В рамках данного технического решения обучающая выборка содержит электронную историю болезни пациентов, сгруппированных по пациенту. Электронная история болезни пациента, используемая для формирования обучающей выборки, содержит записи пациентов, включающих, по меньшей мере, следующие данные: дату добавления записи, коды диагнозов, симптомов, процедур и лекарств, текстовое описание истории болезни на естественном языке, ассоциированные с историей болезни биомедицинские

15 изображения, результаты исследований и анализов пациентов.

В качестве иллюстративного примера приведем следующий фрагмент из истории болезни пациента:

20 01.01.2001

Пациент обратился с жалобами на резкое повышение температуры и кашель с мокротой. Пациент был отправлен на рентген груди и получил жаропонижающе <ЖАРОПОНИЖАЮЩЕЕ> в дозировке <ДОЗИРОВКА>.

25 02.01.2001

Подтверждён диагноз пневмония. Назначено следующее лечение: <ОПИСАНИЕ ЛЕЧЕНИЯ НА ЕСТЕСТВЕННОМ ЯЗЫКЕ>

Используемые форматы представления данных могут варьироваться и меняться в зависимости от используемых технологий. Описанные форматы не являются единственно возможными и описаны для лучшего понимания принципов заложенных в данном

30 техническом решении.

Электронная история болезни пациента может быть представлена в формате openEHR, HL7 и т.д. Выбор формата и стандарта не влияет на сущность технического решения.

В одном из вариантов реализации запись истории болезни представляет собой набор полей, содержащих по меньшей мере параметрами, описывающие:

- 35
- состояние пациента;
 - способы лечения пациента (методики, способы их применения, характеристики);
 - средствами, используемыми при лечении пациента (препараты, дозировки и т.п.);
 - результаты анализов и т.д;

и метаданные, связывающих описанные параметры с параметрами из других записей

40 В случае, если записи в обучающей выборке не сгруппированы по пациенту, после получения данных производят их группировку используя известные алгоритмы или функции (Например, любые сортировки и выборки, известные из уровня техники, в том числе при выборке данных из баз данных использование команд 'GROUP BY', 'ORDER BY' в SQL-запросах).

45 Данные истории болезни могут быть представлены в текстовом, табличном формате, в виде временных рядов, изображений, видео, геномных данных, сигналов, но не ограничиваясь. Данные также могут быть представлены в структурированном и не структурированном виде.

Дата добавления записи может хранить только дату, дату и время, временную отметку, при этом упомянутые записи могут содержать указанные временные объекты, как в абсолютном виде, так и в относительном (относительно временных объектов из других записей).

5 Коды диагнозов, симптомов, процедур, лекарств могут быть представлены в формате МКБ (Например, МКБ-10), SNOMED-CT, CCS (Clinical Classifications Software) или т.д. Выбор формата не влияет на сущность данного технического решения.

Результаты анализов могут быть представлены в табличном виде.

10 Текстовое описание истории болезни может быть представлено в структурированном и не структурированном виде (описание на естественном языке).

Биомедицинские изображения могут быть представлены в виде изображения (jpg, png, tiff и другие графические форматы), видео (avi, mpg, mov, mkv и другие видео-форматы), 3D фото, 3D видео, 3D моделей (obj, max, dwg и т.д.). В виде биомедицинских изображений могут быть представлены результаты ЭКГ, ЭЭГ, МРТ, УЗИ, биопсии, 15 цитологических исследований, рентгена, маммографии и т.д.

Данные РНК-секвенирования могут быть представлены в формате TDF (tiled data format), неиндексированных форматах такие как GFF, BED и WIG, индексированных форматах таких как BAM и Goby, а также в форматах bigWig и bigBed.

20 Описанные выше форматы отражают какое минимум программное обеспечение предназначено для работы с упомянутыми выше данными (создание, модификация и т.д.).

После того, как обучающая выборка сформирована и получена на сервере, сервер производит предварительную обработку данных 220, содержащихся в истории болезней пациентов, выбранных из обучающей выборки.

25 Предварительная обработка данных доменно-специфична и зависит от типа данных и источника данных.

Для каждого типа данных и/или источника данных задаются специализированные обработчики. В случае, когда для типа данных и/или источника не предусмотрен обработчик или в нем нет необходимости, то применяется пустой обработчик или 30 осуществляется пропуск обработчика для данного типа данных.

В одном из вариантов реализации тип данных определяется на основе метаданных, задаваемых для по меньшей мере одного типа поля данных в электронной записи истории болезни пациента.

35 Например, в dicom в метаданных указывается модальность данных в явном виде и модальность данных трактуется согласно внутреннему определению dicom стандарта.

В одном из вариантов реализации тип данных может определяться при помощи сигнатур. При этом на сервере или внешнем источнике есть база данных сигнатур при помощи которых происходит определение типа данных в записи.

40 Например, наличие последовательности байт "GIF89a" в начале данных (поля или файла) обозначает, что это растровое изображение в формате GIF, а наличие байт 'BM' означает, что это растровое изображение в формате BMP.

В одном из вариантов реализации тип данных может определяться на основании информации, содержащейся в записи, с использованием заранее заданных правил.

45 Например, тип данных изображений (Bitmap, Icon), мультимедиа данных (видео, звук), хранящихся в ресурсах исполнимых файлов (PE-файл, .exe) определяется на основании анализа структуры секции ресурсов упомянутого исполнимого файла.

В одном из вариантов реализации данные одного типа могут быть конвертированы в данные другого вида (видео - в совокупность изображений и наоборот, 3d объект - в

изображение проекций упомянутого объекта и наоборот, и т.д.).

Например, для КТ снимков может задаваться обработчик, производящий их трансформацию в серию растровых изображений с возможной нормализацией, если известны параметры устройства, на котором сделан снимок.

5 Еще в одном примере для текста может задаваться обработчик, который производит стандартные для NLP трансформации текста (в основном это приведение к нижнему регистру, замена чисел, удаление стоп-слов и предлогов, стэмминг).

10 Еще в одном примере для текста на естественном языке может задаваться обработчик, который формирует из текста последовательность медицинских фактов при помощи его отображения (англ. mapping - мапинг) на термины медицинской онтологии и/или словаря медицинских терминов.

15 В одном из вариантов реализации для анализа текста на естественном языке могут применяться известные из уровня техники алгоритмы по меньшей мере лексического анализа и синтаксического анализа на основании которых из текста выделяются лексемы и объединяются в объекты, представляющие собой последовательность медицинских фактов.

В одном из вариантов реализации при отображении текста каждый медицинский факт аннотируется (помечается) датой и/или временем, соответствующей дате и/или времени текущей записи из истории болезни.

20 Например, если обработчик обрабатывает поле содержащее текст на естественном языке из записи пациента, имеющей дату 20.01.2017, то все медицинские факты будут аннотированы (помечены) датой 20.01.2017.

25 В одном из вариантов реализации для анализа текста на естественном языке могут применяться известные из уровня техники алгоритмы по меньшей мере лексического анализа и синтаксического анализа на основании которых из текста выделяются лексемы и объединяются в объекты, представляющие собой последовательность медицинских фактов.

30 Еще в одном из вариантов реализации для анализа текста используется предварительно обученная (одним из методов машинного обучения) модель распознавания текста, в результате работы которой формируется совокупность медицинских фактов. При этом упомянутая модель может быть переобучена (с использованием методов обучения с учителем), в случае если сформированные медицинские факты не будут удовлетворять заранее заданным критериям (например, при анализе результатов специалистом).

35 В одном из вариантов реализации обработчика для естественного языка, обработчик ищет каждое слово (после предобработки) из текста в онтологии или словаре. Если слово найдено в онтологии или словаре, то обработчик сохраняет соответствующее ему понятие онтологии или слово из словаря, при этом слова, не найденные в онтологии или словаре, отбрасываются.

40 В одном из вариантов реализации могут использоваться более сложные правила (процедуры) отображения текста на естественном языке в последовательность фактов.

Например, для некоторых понятий могут задаваться дополнительные шаблоны (регулярные выражения), позволяющие извлечь связанные понятия и/или величины.

45 В одном из вариантов реализации онтологии и/или словари располагаются локально на сервере.

Еще в одном из вариантов реализации сервер может получать онтологии и словари из внешних источников,

Например, через *Ontology Lookup Service*, который предоставляет интерфейс веб-

сервиса для запроса многих онтологий из одного места с унифицированным форматом вывода данных.

В одном из вариантов реализации в качестве источника знаний, вместо онтологий или словарей может использоваться любой источник медицинских данных из которого можно сформировать лес знаний (множество ациклических направленных графов знаний). К таким источникам относятся, в частности, медицинские схемы-гайдлайны и т.п.

Еще в одном из вариантов реализации в качестве источника знаний могут выступать медицинские профильные статьи и/или учебники. При этом предварительно найденные статьи обрабатываются известными из уровня техники методами распознавания текстов (с помощью описанного выше лексического и синтаксического анализов, использования обученных моделей распознавания текстов и т.д.).

Еще в одном из вариантов реализации в качестве источника знаний используются открытые биомедицинские онтологии (ОБО, англ. Open Biomedical Ontologies).

В одном из вариантов реализации обработчик производит нормализацию данных (под каждый тип данных могут использоваться свои правила нормализации).

Например, значений измерений определенных показателей крови, оформленных в виде таблиц, обработчик может производить нормализацию упомянутых значений (англ. feature scaling), при этом параметры такого преобразования рассчитываются на обучающей выборке. В частности, вычисляются выборочные среднее μ и дисперсия σ^2 , при этом

$$\tilde{x} = \frac{x - \mu}{\sigma}$$

Еще в одном примере для изображения, где значение каждого пикселя соответствует значению измеренной плотности среды для рентгеновского излучения в Хаундсфилдах обработчик может производить отображение шкалы Хаундсфилда в диапазон [-1, 1] (нормализовать все целочисленные значения в диапазоне [0..255] для черно-белых изображений к действительным значениям в диапазоне [-1.0..1.0]). В частности, нормализация может быть описана формулой:

$$x' = \frac{x - x_{min}}{x_{max} - x_{min}}$$

$$y = y_{min} + (y_{max} - y_{min}) \times x'$$

где

x - нормализуемое значение в пространстве значений $\{X\}$;

x_{min} - минимальное значение в пространстве значений $\{X\}$;

x_{max} - максимальное значение в пространстве значений $\{X\}$;

x' - нормализованное значение в пространстве значений $\{X\}$;

y_{min} - минимальное значение в пространстве значений $\{Y\}$;

y_{max} - максимальное значение в пространстве значений $\{Y\}$;

y' - нормализованное значение в пространстве значений $\{Y\}$;

Еще в одном примере данные из таблицы, содержащей измерения определенных показателей крови, подвергаются препроцессингу - нормализации данных (приведение каждого из параметров к zero mean unit variance, параметры такого преобразования рассчитываются на обучающей выборке).

Еще в одном примере данные, представленные в виде изображения в RGB формате, полученные с микроскопа подвергаются постпроцессингу - бинаризации от вероятности

к классу. Если вероятность патологии по мнению модели больше заданного порога, то изображение помечается как содержащее патологию, иначе как не содержащее.

В одном из вариантов реализации обработчик производит фильтрацию шума или шумопонижение анализируемых данных (процесс устранения шумов из полезного сигнала с целью повышения его субъективного качества или для уменьшения уровня ошибок в каналах передачи и системах хранения цифровых данных). В частности, при обработке изображений может применяться один из методов пространственного шумоподавления (адаптивная фильтрация, медианная фильтрация, математическая морфология, методы на основе дискретного вейвлет-преобразования и т.д.), для видео - один из методов временного, пространственного или пространственно-временного шумоподавления.

Рассмотрим пример, проиллюстрированный на Фиг. 3:

Пусть имеется обучающая выборка 310, состоящая из записей пациентов. Для каждой записи выборки сервер производит предварительную обработку данных, выбранных из полей записи. Так, например, сервер извлекает из обучающей выборки 310 одну запись 301 из истории болезни пациента, определяет состав полей и/или типы данных содержащихся в записи. В данном примере запись 301 содержит дату, описание на естественном языке, снимки КТ, анализ крови. Далее сервер для каждого поля записи производит обработку данных при помощи соответствующего обработчика из пула обработчиков 320 (обработчики 3201..32N). Так, например, дата может быть обработана пустым обработчиком 3201, текст на естественном языке - обработчиком 3202, осуществляющим стандартную для NLP обработку текста, снимки КТ - обработчиком КТ 3203, анализ крови - обработчиком 3204, производящим нормализацию данных. После обработки запись 301 пациента содержит обработанные данные 301*, где символ '*' рядом с полем означает, что в нем содержатся измененные записи (отличающиеся от первоначальных).

В одном из вариантов реализации обработчик формируется с использованием одним из скриптовых языков (языков сценариев), так и в виде плагинов, библиотек (в том числе, представляющих собой исполнимые файлы, типа PE, например, dll).

В одном из вариантов реализации на сервере встроен набор процедур для элементарных действий над типами данных. Комбинирование этих процедур в нужном для пользователя порядке позволяет создавать обработчики самостоятельно. Создание обработчиков в таком случае происходит при помощи встроенной поддержки скриптовых языков или через интерфейс платформы, позволяющий создавать такие обработчики.

После того, как сервер произвел необходимую предварительную обработку данных, сервер преобразовывает 230 обработанные данные в последовательность медицинских фактов по каждому пациенту с использованием медицинских онтологий.

Вся история болезни преобразуется сервером в последовательность медицинских фактов о пациенте. Факты могут содержать дополнительную информацию, например, биомедицинское изображение, ЭКГ, результаты анализов и т.д.

В качестве иллюстративного примера рассмотрим две записи из электронной истории болезни и их результат преобразования в последовательность медицинских терминов.

Запись до преобразования:

45

01.01.2001
Пациент обратился с жалобами на резкое повышение температуры и кашель с мокротой. Пациент был отправлен на рентген груди и получил жаропонижающее <ЖАРОПНИЖАЮЩЕЕ> в дозировке <ДОЗИРОВКА>.
01.01.2001
<Изображение рентгена груди>
02.01.2001

Подтвержден диагноз пневмония. Назначено следующее лечение: <ОПИСАНИЕ ЛЕЧЕНИЯ НА ЕСТЕСТВЕННОМ ЯЗЫКЕ>”

Полученные после преобразования медицинские факты.

“01.01.2001

<ТЕМПЕРАТУРА>

<КАШЕЛЬ>

<РЕНТГЕН ГРУДИ. Изображение>

<ЖАРОПОНИЖАЮЩЕЕ, ДОЗИРОВКА>

02.01.2001

<ПНЕВМОНИЯ>

<ПРЕПАРАТ1, ДОЗИРОВКА>

<ПРЕПАРАТ2, ДОЗИРОВКА>

<ПРОЦЕДУРА1>

После преобразования истории болезни в набор медицинских фактов, сервер производит автоматическую разметку полученной последовательности медицинских фактов 240 по каждому пациенту, используя извлеченные из истории болезни пациента диагнозы или другие интересующие факты. В случае, если данные размечены, то этот шаг пропускается сервером.

В одном из вариантов реализации интересующие факты задаются пользователем сервера или собираются у пользователей данного технического решения (например, у врача).

Например, в качестве интересующих фактов могут выступать списки включающих и исключаяющих критериев для клинических испытаний, т.е. списки критериев, которым должен соответствовать человек, чтобы быть включенным в клинику (включающие списки), или наоборот быть исключенным из клиники или не быть допущенным в клинику (исключающие списки).

Еще в одном примере включающим критерием может быть заболевание раком печени с опухолью размером не больше 5 мм.

Еще в одном примере исключаяющим критерием может быть курение, возраст пациента выше 65 лет.

В одном из вариантов реализации интересующие факты извлекаются из внешних источников.

Например, интересующие факты могут извлекаться из медицинских информационных систем.

Далее сервер упорядочивает и группирует факты по осмотрам по времени. Такая группировка необходима для того, чтобы рассматривать группу фактов внутри одного осмотра одновременно.

В одном из вариантов реализации анализы могут относиться к приему, на котором они были назначены, или выделяются в отдельную сущность (отдельный осмотр).

В одном из вариантов реализации КТ, МРТ, гистологию относят к отдельному осмотру.

Еще в одном из вариантов реализации к отдельным осмотрам относят по меньшей мере все методы исследования, содержащие необработанные данные (не отчет врача, а непосредственный результат в виде изображения, видео, временных меток). Если доступен только отчет врача или сам факт прохождения исследования, то такие данные

5 рассматриваются как часть осмотра.

Для такой группировки по осмотрам сервер использует информацию о времени и/или дате, связанные с каждым фактом.

Далее сервер формируются пары {множество фактов, диагноз} или {множество фактов, интересующий факт} на основании группировки по осмотрам.

10 В одном из вариантов реализации пары формируются простым перебором.

Далее сервер подготавливает обучающую выборку для каждой из модальности данных. Как упоминалось ранее, в качестве модальности данных могут выступать данные гистологии, рентгена, КТ, маммографии и т.д.

15 Например, для формирования обучающей выборки для модальности КТ сервер отбирает записи, содержащие КТ:

```
{<КТ, изображение>, <ДИАГНОЗ_п>}, {<КТ, изображение>, <ДИАГНОЗ_к>},
{<КТ, изображение>, <ДИАГНОЗ_к>}, ...
Например, для концептов (терминов), сервер отбирает следующие записи
{<ТЕМПЕРАТУРА>, <ПНЕВМОНИЯ>}
{<ТЕМПЕРАТУРА>, <КАШЕЛЬ>, <ПНЕВМОНИЯ>}
{<ТЕМПЕРАТУРА>, <КАШЕЛЬ>, <РЕНТГЕН ГРУДИ>, <ПНЕВМОНИЯ>}
{<ТЕМПЕРАТУРА>, <КАШЕЛЬ>, <РЕНТГЕН ГРУДИ>, <ЖАРОПонижающее>},
<ПНЕВМОНИЯ>}
```

20 Затем, после формирования обучающих выборок для каждой модальности, сервер производит обучение первичных репрезентаций 250 независимо для каждой из модальностей.

Для каждой модальности на сервере задается модель (группа моделей), для обучения прогнозирования выявленных в данной обучающей выборке диагнозов, присутствующих

25 в данных модальностях.

В одном из вариантов реализации в качестве модели могут выступать такие алгоритмы машинного обучения, как:

- линейная регрессии;
- логистическая регрессия;
- 30 • алгоритм к ближайших соседей;
- случайный лес;
- градиентный бустинг на деревьях;
- байесовы классификаторы;
- 35 • глубокие нейронные сети (полносвязные, сверточные, рекуррентные, их комбинации).

Еще в одном из вариантов реализации упомянутая модель удовлетворяет требованию - соответствие модальности, с которым эта модель будет работать.

Например, для изображений используются сверточные сети.

40 В одном из вариантов реализации для каждой модальности задается несколько моделей.

В одном из вариантов реализации модальности группируют в кластеры (например, вся гистология, рентген, маммография и др.), имеющие общую архитектуру моделей (одно параметрическое семейство) и обучающиеся совместно, при этом, каждая модель из кластера имеет разные комплекты весов.

45 В одном из вариантов реализации для каждой модальности формируется набор параметрических семейств моделей. Параметрическое семейство означает, что есть общий вид моделей с некоторым набором параметров, определение которых однозначно задает модель.

Например, если в качестве модели используется нейронная сеть, то один из примеров параметрического семейства это многослойный перцептрон, а параметрами будут количество слоев и количество нейронов в слоях. Другой пример - любая нейросеть с фиксированной архитектурой, которая порождает параметрическое семейство (к примеру семейство, предназначенное для сегментации изображений, для классификации изображений и т.д.).

В рамках данного технического решения предполагается использование следующих основных параметрических семейств:

- сверточные нейронные сети для работы с изображениями, видео, сигналами;
- рекуррентные нейронные сети для работы с последовательностями фактов в истории болезни пациента и для построения прогнозных моделей, для обработки неструктурированной текстовой информации;

- Байесов подход и деревья решений для работы с табличными данными.

Разберем подробнее на примере работы с изображениями.

В работе с изображениями встречаются следующие основные задачи: классификация изображения (поставить в соответствие каждому изображению один или несколько классов, или меток), сегментация изображений (поставить одну или несколько меток для каждого пикселя изображения), локализация объектов интереса (построить обрамляющий прямоугольник, внутри которого находится объект интереса, для каждого объекта на изображении). Для каждой из этих задач задается архитектура, решающая данную задачу. Отличия у модальностей в основном в размере входного изображения и в количестве целевых меток/объектов интереса. За основу каждой такой модели берется Dense-net, концепт архитектуры которой представлен на фиг. 5.

Идея данной архитектуры в использовании дополнительных путей для движения информации внутри модели, что позволяет эффективно обучать даже очень большие модели с большим количеством сверточных слоев. При заданной модальности, размере входного изображения и количества классов, такая модель образует параметрическое семейство, и веса нейронной сети как раз и является параметрами семейства. Они определяются в процессе обучения, во время которого модели предъявляются изображения и целевые метки, и нейронная сеть изменяет свои веса так, чтобы ее отклик совпадал с тем, что содержится в разметке обучающей выборки (так называемые целевые значения или целевой отклик).

Далее для каждой модальности сервер производит поиск параметров семейства, дающих оптимальный результат на данной обучающей выборке.

В одном из вариантов реализации для поиска параметров семейства, дающих оптимальный результат, используется по меньшей мере:

- метод Монте-Карло;
- Байесова оптимизация.

В одном из вариантов реализации для оценки моделей использует кросс-валидация (скользящий контроль или кросс-проверка), на основании которой выбирается лучшая по показателям модель для данной обучающей выборки для данной модальности.

Процедура скользящего контроля осуществляется следующим образом. Выборка X^L разбивается N различными способами на две непересекающиеся подвыборки:

$$X^L = X_n^m \cup X_n^k,$$

где

X_n^m - обучающая подвыборка длины m ,

X_n^k - контрольная подвыборка длины $k=L-m$,
 $n=1..N$ - номер разбиения.

Для каждого разбиения n строится алгоритм

$$5 \quad a_n = \mu(X_n^m)$$

и вычисляется значение функционала качества

$$Q_n = Q(a_n, X_n^k)$$

Среднее арифметическое значений Q_n по всем разбиениям называется оценкой
 10 скользящего контроля:

$$CV(\mu, X^L) = \frac{1}{N} \sum_{n=1}^N Q(\mu(X_n^m), X_n^k).$$

Именно на основании оценки скользящего контроля и выбирается лучшая модель.

В одном из вариантов реализации модели добавляются по мере сбора данных и
 15 добавления новых источников данных о пациентах.

Например, в медицинской информационной системе стал доступен новый тип исследований (например, ЭКГ)- Далее для модальности ЭКГ формируется обучающая выборка, затем на этой обучающей выборке обучается нейросетевая (фиг. 4) модель (группа моделей), из которой получается репрезентация этой модальности.

20 После того, как модели обучены, сервер формирует первичные векторные репрезентации для каждой модальности. Для этого на вход каждой модели, обученной для данной модальности, сервер подает ранее предобработанные данные пациентов, определяет выходные значения модели и значения весов последнего скрытого слоя данной модели для каждой записи. Значения весов последнего скрытого слоя далее
 25 будут использоваться сервером в качестве первичных векторных репрезентаций, которые представляют собой отображение модальности в вектор фиксированного размера, определяемого моделью. В результате сервером формируется новый набор данных, представляющий собой трансформацию исходного.

У каждой модальности есть своя размерность вектора, например, если модальность
 30 - m_u , а размерность общего пространства - n , то будет строиться отображение:

$$R^{m_u} \rightarrow R^n$$

Например, пусть

$$x \in R^{m_u}, y \in R^n,$$

35 тогда отображение:

$$y = Ax + b,$$

где

$$A \in R^{(n, m_u)}, b \in R^n$$

40 Еще в одном примере

$$y = f(Ax + b),$$

где

f - нелинейная функция (например, ReLU, sigmoid, и т.д.).

45 То есть A есть матрица размером n, m_u , а b есть вектор размера n

В одном из вариантов реализации векторная репрезентация текста строится в пространстве, в которое отображаются первичные векторные репрезентации всех остальных модальностей, т.е. первичная векторная репрезентация текстовой

модальности отображается в пространство общих репрезентаций тождественным отображением.

В случае использование не нейросетевой модели возможны два сценария:

- в качестве репрезентации данных принимается выход модели как таковой;
 - нейросетевая модель представляет собой классификатор, отображающий входные данные в некоторый набор фактов, для которых уже есть векторные представления.
- Поскольку гарантируется, что любая модель будет порождать вектор вероятности наличия факта, то векторная репрезентация будет взвешенная по вероятности сумма вектор-репрезентаций фактов.

Например, строится модель

$$f(x, \xi): X \rightarrow Y,$$

где

X - множество признаков,

$Y = \{y_i, i = 1, \dots, n\}$ - множество целевых фактов

ξ - параметры модели.

Без ограничения общности можно переформулировать задачу следующим образом:

$$P = f(x, \xi): X \rightarrow R^n,$$

при этом

$$P = (p_1, p_2, \dots, p_n); \sum_{i=1}^n p_i = 1, p_i \geq 0.$$

При таких ограничениях на модель p_i можно трактовать как вероятность наличия факта y_i у пациента (либо другие варианты, в зависимости от решаемой задачи, например, появление факта с горизонтом год и так далее).

Имея обучающую выборку $\{x_j, t_j\}, j = 1, \dots, N$, можно определить параметры модели ξ . Обозначим найденные в процессе обучения параметры модели как $\hat{\xi}$. Также, поскольку каждый из y_i представляет собой медицинский факт, то ему соответствует векторная репрезентация V_i .

Тогда, для нового случая получаем следующее: строим соответствующий этому случаю вектор входных признаков \bar{x} ; получаем соответствующий ему вектор вероятностей $\bar{p} = f(\bar{x}, \hat{\xi})$; векторная репрезентация данной модальности в данном случае будет строиться следующим образом:

$$\bar{V} = \sum_{i=1}^n \bar{p}_i V_i.$$

Сервер формирует первичные векторные репрезентации для каждой из модальностей в результате чего получается набор векторных репрезентаций для медицинских фактов и терминов (диагнозов, симптомов, процедур и медикаментов) и модели для отображения первичных векторных репрезентаций в пространство совместных репрезентаций.

В одном из вариантов реализации сервер дополнительно предобучает векторные репрезентации медицинских терминов (концептов), например,

- при наличии дополнительного источника данных в виде большого корпуса медицинской литературы;
- при наличии альтернативного обучающего корпуса, который был собран независимо от текущего.

Сервер осуществляет предобучение медицинских терминов (концептов) с

использованием дистрибутивной семантики и векторного представления слов.

Каждому слову присваивается свой контекстный вектор. Множество векторов формирует словесное векторное пространство.

В одном из вариантов реализации предобучение медицинских терминов (концептов) осуществляется при помощи программного инструмента анализа семантики естественных языков Word2vec с использованием онтологии для регуляризации.

Регуляризация в статистике, машинном обучении, теории обратных задач - метод добавления некоторой дополнительной информации к условию с целью решить некорректно поставленную задачу или предотвратить переобучение. Эта информация часто имеет вид штрафа за сложность модели.

Например, это могут быть:

- ограничения гладкости результирующей функции;
- ограничения по норме векторного пространства;
- регуляризация на весах и на активациях нейронов;
- известные из уровня техники методы регуляризации.

В данной технической решении используются известные из уровня техники основные и общепринятые для машинного и глубокого обучения способы регуляризации. Допустим, что E - функция ошибки, минимизируемая в процессе обучения, W - веса модели, A - активация всех нейронов скрытых слоев (если речь ведется о нейронной сети). Тогда, одна из наиболее широко используемых техник регуляризации под названием L_1 (L_2) регуляризация может быть описан следующим образом: Вместо минимизации E решается задача минимизации

$$E + \alpha L_1(W) \xrightarrow{W} \min \text{ (} L_1 \text{ регуляризация весов),}$$

$$E + \alpha L_2(W) \xrightarrow{W} \min \text{ (} L_2 \text{ регуляризация весов),}$$

$$E + \alpha L_1(A) \xrightarrow{W} \min \text{ (} L_1 \text{ регуляризация активаций),}$$

$$E + \alpha L_2(A) \xrightarrow{W} \min \text{ (} L_2 \text{ регуляризация активаций),}$$

где

$$L_p(x) = (\sum_{i=1}^n |x|^p)^{\frac{1}{p}} \text{ - } L_p\text{-норма.}$$

Возможны также различные варианты приведенных случаев. Данные регуляризующие слагаемые (термы) накладывают дополнительные (нежесткие) ограничения. Т.е. не заданные в виде явной системы уравнений и/или неравенств, порождающих множество $\tilde{W} \subset R^n$ допустимых весов модели) ограничения на возможные веса модели, что позволяет избежать переобучения. Также, помимо L_1/L_2 регуляризации могут быть использованы:

- ранний останов:

суть данного метода в том, чтобы из обучающей выборки выделить небольшую тестовую выборку, которая не участвует явно в процессе обучения, но используется для измерения ошибки модели в процессе обучения. Как только ошибка на этой тестовой выборке начинает расти, обучение прекращается.

- синтетическое увеличение обучающей выборки (англ. data augmentation):

суть данного подхода в том, что к каждому примеру обучающей выборки с некоторой вероятностью применяется преобразование, не меняющий желаемый отклик или позволяющее, применив аналогичное преобразование, получить новый ожидаемый

отклик, который будет корректным. Например, классифицируя рентгеновское изображение грудной клетки на предмет наличия или отсутствия признаков пневмонии, к входному изображению можно применить зеркальное отображение относительно вертикальной оси, т.к. очевидно это не изменит целевую метку.

5 • могут быть явно наложены ограничения на параметры модели через ограничение на значение норм вектора весов модели: $L_1(W) < \gamma$ или $L_2(W) < \gamma$.

• могут применяться и другие методы регуляризации, широко используемые в машинном и глубоком обучении.

10 Иллюстративно опишем работу инструмента word2vec: на вход word2vec принимает большой текстовый корпус и сопоставляет каждому слову вектор, выдавая координаты слов на выходе. Сначала он создает словарь, «обучаясь» на входных текстовых данных, а затем вычисляет векторное представление слов. Векторное представление основывается на контекстной близости: слова, встречающиеся в тексте рядом с одинаковыми словами (а, следовательно, имеющие схожий смысл), в векторном
15 представлении будут иметь близкие координаты векторов-слов. Полученные векторы-слова (например, фиг. 9а для бронхита и фиг. 9б для ревматита) могут быть использованы для обработки естественного языка и машинного обучения.

В word2vec существуют два основных алгоритма обучения: CBOW (Continuous Bag of Words) и Skip-gram. CBOW («непрерывный мешок со словами») - модельная
20 архитектура, которая предсказывает текущее слово, исходя из окружающего его контекста. Архитектура типа Skip-gram действует иначе: она использует текущее слово, чтобы предугадывать окружающие его слова. Порядок слов контекста не оказывает влияния на результат ни в одном из этих алгоритмов.

Получаемые на выходе координатные представления векторов-слов позволяют
25 вычислять «семантическое расстояние» между словами. Основываясь на контекстной близости этих слов, технология word2vec делает свои предсказания.

В одном из вариантов реализации при использовании онтологии как регуляризации (ограничения на структуру пространства) используется attention.

30 В одном из вариантов реализации при использовании онтологии для регуляризации (ограничения на структуру пространства) используется многоуровневая (иерархическая) функция ошибки, опирающаяся на соотношения между терминами из онтологии.

Онтология используется в частном случае в виде графа знаний, который задает иерархию терминов и их категорий. Это позволяет заранее упорядочить пространство векторного
35 представления, так как очевидно, что близость в графе знаний должна означать близость в векторном пространстве между терминами или категориями. Используя это, можно наложить штраф на обучение векторных презентаций. В дальнейшем минимизируется штраф вместе с основной функцией ошибки. Обозначим за c вектор текущего термина. За q обозначим бинарную меру близости между двумя терминами c_1 и c_2 по онтологии. Если $q(c_1, c_2) = 0$, то термины можно считать близкими, если $q(c_1, c_2) = 1$, то далекими.
40

Тогда функция ошибки на онтологии может быть задана:

$$OD(c_i) = \sum_{c_j: q(c_i, c_j) = 0} \|c_i - c_j\| - \sum_{c_j: q(c_i, c_j) = 1} \|c_i - c_j\|$$

45 Теперь, в процессе обучения векторных репрезентаций $OD(c)$ может использоваться по аналогии с L_1/L_2 регуляризацией. Использование регуляризации с помощью онтологии позволяет улучшить качество модели, не прибегая к расширению обучающей выборки. За счет разумного ограничения на пространство репрезентаций, накладываемые

регуляризацией, повышается качество модели, что позволяет в частности избежать переобучения, стать алгоритму более устойчивым по отношению к выбросам и ошибкам в обучающем наборе. Стандартные методы классической регуляризации также накладывают ограничения на пространство репрезентаций, но они только сужают варианты, в отличие от регуляризации на онтологии, которая накладывает ограничения на пространство репрезентаций, основанные на внешней информации о доменной области.

В одном из вариантов реализации онтология, используемая для регуляризации, представляет собой внешний относительно системы параметр, может задаваться заранее и зависеть от соответствующей системы кодов заболеваний (например, Icd9/10, МКБ-10 и т.д.).

Для каждой нейросетевой модели, полученной на данном шаге, извлекается первичная векторная репрезентация как выход скрытого слоя. Упомянутая модель позволяет отобразить входные данные заданной модальности в первичное векторное представление. Это требует простых манипуляций с обученной моделью, по сути сводящейся к удалению из модели выходного слоя.

После обучения и получения первичных репрезентаций сервер осуществляет обучение совместных репрезентаций 260 (проиллюстрировано на фиг. 10) (англ. coordinated multimodal machine learning, подробнее "Multimodal Machine Learning: A Survey and Taxonomy", Tadas Baltrusaitis, Chaitanya Ahuja, and Louis-Philippe Morency).

Для того, чтобы использовать в процессе обучения совместных репрезентаций нетекстовые данные, например, медицинские изображения, необходимо обучить модель отображения из пространства этой модальности в общее векторное пространство. Для этого используется первичная векторизация модальности и обучаемая функция представления из первичной векторной репрезентации в общую, при этом в качестве обучаемой функции может выступать описанная выше модель (модель, обученная для отображения из пространства заданной модальности в общее векторное пространство).

Например, если модель занимается исключительно классификацией изображения, то за последним сверточным слоем может следовать несколько скрытых полносвязных слоев. В таком случае берется выход именно последнего скрытого сверточного, а не полносвязного слоя.

Когда же в истории болезни встречается нетекстовая модальность, например, КТ скан, берется его первичная векторная репрезентация, обрабатывается с использованием многослойного перцептрона, при этом выход этого перцептрона считается вектором-репрезентацией данного изображения и на нем считается косинусное расстояние с его соседями.

Далее используется skip-gram, но, когда дело доходит до нетекстовых модальностей (например, медицинские изображения), в качестве их векторных представлений, и используются выход функции для этой модальности, при этом на вход skip-gram передается корпус из последовательностей медицинских фактов, извлеченный из историй болезни или медицинских текстов.

Затем, после обучения совместных репрезентаций сервер производит обучение финальных моделей и параметров агрегации 270.

Агрегация представляет собой получение из набора векторов, где каждый вектор представляет медицинский факт из истории болезни выбранного пациента единого вектора.

Каждому факту присваивается вес, получаемый в процессе обучения. Формируется набор весов, которые используются в процессе получения прогноза/диагноза для

конкретного пациента - параметры агрегации. Затем определяют взвешенную сумму - домножая каждый вектор в истории болезни пациента на соответствующий ему вес и полученные вектора суммируются. Обычно, при агрегации вектор-репрезентаций используется прямая сумма векторов $c_{ag} = \sum_{i=1}^k c_i$, где c_{ag} агрегированное представление пациента, c_i - векторные репрезентации фактов в истории данного пациента. Однако, такой вариант агрегации не всегда может быть оптимальным, так как каждый из фактов может иметь разный вес точки зрения принятия решения для каждой из нозологии или же для этого пациента. Поэтому предлагается использовать в качестве агрегации следующий подход $c_{ag} = \sum_{i=1}^k a_i c_i$, где a_i есть скаляр, $\sum_{i=1}^k a_i = 1$. Каждый их a_i может быть как явным параметром модели и определяться в процессе обучения, так и представлять из себя функцию вида $a_i = f(i, c_1, c_2, \dots, c_k, \psi)$, где ψ есть параметры этой функции, которые определяются в процессе обучения со всеми остальными весами модели.

Строится граф вычислений, где веса присутствуют как параметры. Далее параметры графа оптимизируются под текущий набор данных методом градиентного спуска. Получаемый в результате набор весов является обучаемым, то есть модифицируется вместе с остальными весами модели в процессе обучения. Веса определяют конкретную функцию из параметрического семейства, которая из нескольких входных векторов формирует один выходной.

Все вышесказанное можно резюмировать следующим образом: Обучение классификатора под группу диагнозов осуществляется на основе графов, обучающая выборка генерируется из доступных EHR в автоматическом режиме на базе NLP-техник (извлечение фактов + расположение их во временном порядке, дальше из них генерируются пары «факты»-«диагнозы»). Выбор классификатора детерминирован возможностью осуществлять работу с векторными некатегориальными признаками и в данном способе это многослойные полносвязные нейронные сети с остатками.

На Фиг. 7 изображены два пайплайна (pipeline) для ЭКГ и биомедицинских изображений, например, КТ грудной клетки. Данные попадают сначала в блок препроцесса. Блок препроцесса доменно-специфичный и преобразует входные данные к виду, который модель ожидает получить. Преобразованные данные отправляются в соответствующую модальности упомянутых данных модель - например, КТ грудного отдела отправляется в нейросеть, которая и анализирует данное исследование. Модель может выдавать результаты в двух видах (два выхода): это собственно желаемый выход модели (вероятность патологии, карты сегментации и т.д.) и векторную репрезентацию данного конкретного примера. Желаемый выход модели отправляется в модуль постпроцесса, который сопряжен с моделью, и выход этого блока показывается человеку-эксперту, например, или отправляется заказчику в виде отчета или в другом, удобном для него виде.

В центральной схеме изображено векторное пространство медицинских концептов, которое строится на основе skip-gram+регуляризация, но онтологии и каждый концепт отображается в определенную точку этого пространства. Для каждой модели также строится отображение в это пространство медицинских концептов из векторной репрезентации, которая порождена моделью из пайплайна, через функцию отображения.

Далее из этого общего пространства забираются вектора для конкретного пациента и отправляются в финальную модель, где агрегируются в единую модель пациента и по ней ставится диагноз и/или рекомендуется лечение и т.д.

После того, как все необходимые действия выполнены администратор, врач или ной

пользователь добавляет на сервер (отправляют) записи пациента, которые необходимо проанализировать.

Примером такой записи может быть:

5 Мужчина в возрасте 60 лет обратился с жалобами на боли в груди при ходьбе, отдышку. К истории прикреплено описание ЭКГ исследования: На ЭКГ выявлен синусовый ритм, QRS комплекс в норме, выраженная депрессия ST сегмента (1 мм) в отведениях V4-V6. Лечение не назначен. Дифференциальный диагноз отсутствует.

Затем сервер осуществляет предобработку данных, выделение ключевых фактов и преобразование их в медицинские факты, например:

10 <Боли в груди>
<Отдышка>
<QRS комплекс в норме>
<депрессия ST сегмента>

15 Далее сервер отправляет полученный набор фактов на вход существующим моделям и определяет диагноз с наибольшей вероятностью соответствующий предъявленному набору фактов.

Затем сервер получает результаты применения модели. В качестве иллюстративного результата на данном примере сервер выдает следующие результаты модели: 75% соответствия диагнозу «стенокардия», рекомендуемы дополнительные исследования: 20 велоэргометрия, суточное ЭКГ мониторингирование.

Отображение результатов может быть в виде рекомендаций, выделения регионов интереса на медицинских изображениях, в виде отчетов.

В качестве результата применения модели (моделей) может быть анализ и прогноз развития заболеваний.

25 В некоторых вариантах реализации может быть предсказана смертность пациента.

В некоторых вариантах реализации на сервер поступает список запросов. Например, запрос - это одна серия КТ исследования требующая обработки сервером. Сервер осуществляет обработку и, например, на КТ срезе красным кругом обводит (выделяет) 30 потенциальный регион интереса, определенный моделью, при этом, списком приводятся все найденные объемы интереса состоящие из регионов интереса. Регион интереса локализуется на слайсе, объемы интереса строятся как агрегация нескольких регионов интереса в один.

В некоторых вариантах реализации может быть предсказана сердечная недостаточность, заболевания печени и другие болезни (патологии).

35 Ниже приведены экспериментальные данные по использованию данного технического решения в одном из вариантов его реализации.

Для исследования были взяты данные MIMIC-III (Medical Information Mart for Intensive Care III) - свободно доступной базы, содержащей обезличенные данные о здоровье, ассоциированные почти с 40 тысячами пациентов, которые пребывали в отделении 40 интенсивной терапии медицинского центра Бет-Изрейел в интервале с 2001 по 2012 годы.

MIMIC-III содержит информацию о демографии, лабораторных измерениях, процедурах, прописанных лекарствах, смертности и жизненных показателях пациентов, зафиксированных во время нахождения пациента в медицинском центре.

45 Для того, чтобы сравнить между собой разные модели, были обработаны данные с помощью следующего подхода.

Сначала была извлечена из базы информация по диагнозам (в виде ICD9 кодов, которые использовались в исходной базе), прописанным лекарствам (в виде

идентификаторов NDC, или, если они отсутствовали, в виде текстового описания лекарства) и назначенным процедурам (в виде ICD9 кодов процедур) для каждого из пациентов с привязкой к конкретному приему, упорядочили по дате приема.

5 Полученная в результате выборка содержала большое количество пациентов с короткой, но малоинформативной историей посещений лечебного заведения. Можно использовать информацию о таких пациентах, чтобы обучать матрицу весов для установления связей между диагнозами. Однако, при обучении моделей такая информация не будет полезна: событие, произошедшее во время следующего обращения к врачу, нельзя точно узнать для пациентов, обратившихся в клинику только 1 раз.

10 Поэтому для подготовки матриц весов была использована вся информация о пациентах, а для обучения моделей дополнительно обрабатывали данные: проходили по истории пациента скользящим окном длиной в год и рассматривали все посещения клиники, зафиксированные в течение этого года, в качестве независимого набора признаков, при этом исключили из рассмотрения скользящие окна, в которых было
15 меньше 2 посещений.

Из каждой последовательности посещений медицинского центра пациентом в течение года были извлечены все приемы, кроме последнего, и использовали их и связанную с ними информацию для извлечения признаков, передаваемых на вход конкретным моделям. Затем провели разделение всех таких последовательностей на обучающую и
20 тестовую выборки в соотношении 4 к 1.

Дополнительная обработка может зависеть от того, что принимает на вход конкретная модель.

СУБД MIMIC-III спроектирована таким образом, что каждому посещению пациентом медицинского центра может быть сопоставлено несколько диагнозов, лекарств,
25 процедур, при этом их порядок внутри приема неоднозначен. Поэтому если модель принимает на вход упорядоченную последовательность событий, при этом не учитывает время посещения, то при ее обучении диагнозы, лекарства и процедуры внутри одного приема переупорядочивают случайным образом, а затем производят объединение "приемов" в последовательность.

30 Поскольку такая последовательность событий будет иметь разную длину для разных пациентов, и далекие по времени события будут оказывать меньший вклад в предсказание диагнозов, в некоторых вариантах реализации обучают модели на последних N событиях (при меньшем количестве событий в истории пациента - дополняют их нулями). Только для модели рассматривается вся последовательность
35 посещений для каждого окна длиной в 1 год.

Предобучение матрицы весов

При построении некоторых классификаторов были использованы так называемые embedding матрицы, чтобы получить MCV-вектор, или вектор медицинских концептов - сжатое представление истории болезни пациента, представляющее собой вектор
40 конечной длины, элементы которого - вещественные числа.

6. Схема получения векторов медицинских концептов для пациента показана на фиг.

Для его построения были рассмотрены события в истории пациента, упорядоченные по времени. Если встречалось событие определенного типа, то записываем, "1" в позицию
45 вектора, соответствующую событию. Так получаем вектор большой размерности, состоящий из 1 - в тех позициях, которые соответствуют событиям, произошедшим в истории пациента, и 0 - в позициях не произошедшим в истории пациента.

В некоторых моделях эта общая схема модифицируется: например, рассматриваются

случайно переупорядоченные в рамках одного приема (и одной временной метки) события, в вектор событий записывается фиксированное количество последних событий, либо события, соответствующие одному приему, дополнительно умножаются на вес, который определяется тем, насколько давно по времени произошло событие.

5 Такое разреженное представление истории пациента занимает много памяти, обучение модели на основе него требует много времени. Для того, чтобы сократить время и сжать данные, на основе разреженного представления истории болезней строят вектор медицинских концептов для пациента.

10 Для получения сжатого представления в простейшем случае используется так называемая embedding-матрица, на которую домножается разреженный вектор истории пациента. Были рассмотрены несколько матриц:

Word2Vec: в качестве матрицы мы брали матрицу коэффициентов, полученную на основании анализа сопутствующих диагнозов, предписаний и медикаментов, которые встречались в рамках одного посещения пациентом врача. Для обучения мы
15 использовали модель word2vec с механизмом skipgram, чтобы для любого диагноза, медицинской процедуры или прописанного лекарства (соответствующего столбцу embedding-матрицы) получать вектор медицинских концептов определенной длины.

Мы обучали эту весовую матрицу на истории болезни пациента, включающей коды диагнозов, назначенных процедур и прописанных лекарств, чтобы извлечь больше
20 информации о связях между диагнозами.

Ontology embedding: Использовалась информация об онтологии, а именно, для получения сжатых представлений событий использовались коды, которые расположены в вышестоящих узлах дерева диагнозов, выраженного в терминах ICD9-кодов.

Embedding with ICD9 tree: для получения сжатого представления может использоваться
25 нестандартную функцию регуляризации, которая максимизирует расстояние до далеких и минимизирует расстояние до близких объектов в дереве (и одновременно с этим корректирует вектора для родительских узлов в дереве ICD9-кодов). В отличие от имеющихся подходов, в которых происходит обучение весов диагнозов и их родителей в дереве под конкретную задачу, предобучается матрица весов, и уже в обученном
30 варианте она используется для обучения моделей.

Для каждой из поставленных задач могут быть рассмотрены несколько классификаторов:

Basic 1-hot encoding: эта модель была построена на основе логистической регрессии, на вход которой подавался массив из 0 и 1 - разреженное векторное представление
35 истории болезни пациента. Рассматривались только заболевания.

TFI-DF encoding: эта модель была построена на основе логистической регрессии, на вход которой подавался массив, ячейки которого были сопоставлены заболеваниям пациента. Построение проведено аналогично предыдущей модели, за исключением того, что для заболеваний учитывалось, сколько раз они встретились в истории пациента,
40 а затем входные признаки были обработаны алгоритмом TF-IDF, чтобы сопоставить больший вес редко встречающимся в целом, но часто замеченным у конкретного пациента диагнозам.

Следующие несколько моделей использовали похожую архитектуру нейронной сети для классификации, но разные матрицы весов.

45 Word2Vec embeddings: модель использовала матрицу весов для получения сжатого представления истории болезни в виде вектора пациента. В качестве матрицы весов использовалась матрица Word2vec, основанная на механизме skip-gram. Полученные сжатые представления использовались в качестве признаков для логистической

регрессии.

Word2Vec embedding+attention: в модели использована Word2vec-матрица весов для получения сжатого представления вектора пациента внутри модели. Кроме того, была использована нейросетевая архитектура с механизмом внимания.

5 Embedding with ICD9 tree: модель с embedding-матрицей, построенной на основе дерева icd9-кодов. Сжатые представления пациентов, полученные умножением матрицы на вектора пациентов, были использованы для построения модели на основе логистической регрессии.

10 Embedding with ICD9 tree + attention: модель с embedding-матрицей, построенной на основе дерева icd9-кодов, в которой была использована архитектура нейронной сети с механизмом внимания.

Embedding with ICD9 tree + attention + tfidf: модель, которая отличается от предыдущей тем, что дополнительно на вход ей было передано значение, которое возвращала для заданного пациента модель TFI-DF encoding.

15 Choi embedding+attention: модель с embedding-матрицей, построенной на основе сжатых представлений векторов, рассмотренных в работе Choi et all «GRAM: Graph-based Attention Model for Healthcare Representation Learning», с использованием механизма внимания.

Time-based model: Воспроизведен способ построения msv-векторов пациентов.

20 На Фиг. 8 показан пример компьютерной системы общего назначения, на которой может выполняться описываемое техническое решение и которая включает в себя многоцелевое вычислительное устройство в виде компьютера 20 или сервера, включающего в себя процессор 21, системную память 22 и системную шину 23, которая связывает различные системные компоненты, включая системную память с процессором 21.

25 Системная шина 23 может быть любого из различных типов структур шин, включающих шину памяти или контроллер памяти, периферийную шину и локальную шину, использующую любую из множества архитектур шин. Системная память включает постоянное запоминающее устройство (ПЗУ) 24 и оперативное запоминающее устройство (ОЗУ) 25. В ПЗУ 24 хранится базовая система ввода/вывода 26 (БИОС), состоящая из основных подпрограмм, которые помогают обмениваться информацией между элементами внутри компьютера 20, например, в момент запуска.

30 Компьютер 20 также может включать в себя накопитель 27 на жестком диске для чтения с и записи на жесткий диск, накопитель 28 на магнитных дисках для чтения с или записи на съемный диск 29, и накопитель 30 на оптическом диске для чтения с или записи на съемный оптический диск 31 такой, как компакт-диск, цифровой видео-диск и другие оптические средства. Накопитель 27 на жестком диске, накопитель 28 на магнитных дисках и накопитель 30 на оптических дисках соединены с системной шиной 23 посредством, соответственно, интерфейса 32 накопителя на жестком диске, интерфейса 33 накопителя на магнитных дисках и интерфейса 34 оптического накопителя. Накопители и их соответствующие читаемые компьютером средства обеспечивают энергонезависимое хранение читаемых компьютером инструкций, структур данных, программных модулей и других данных для компьютера 20.

45 Хотя описанная здесь типичная конфигурация использует жесткий диск, съемный магнитный диск 29 и съемный оптический диск 31, специалист примет во внимание, что в типичной операционной среде могут также быть использованы другие типы читаемых компьютером средств, которые могут хранить данные, которые доступны с помощью компьютера, такие как магнитные кассеты, карты флеш-памяти, цифровые видеодиски, картриджи Бернулли, оперативные запоминающие устройства (ОЗУ), постоянные

запоминающие устройства (ПЗУ) и т.п.

Различные программные модули, включая операционную систему 35, могут быть сохранены на жестком диске, магнитном диске 29, оптическом диске 31, ПЗУ 24 или ОЗУ 25. Компьютер 20 включает в себя файловую систему 36, связанную с операционной системой 35 или включенную в нее, одно или более программное приложение 37, другие программные модули 38 и программные данные 39. Пользователь может вводить команды и информацию в компьютер 20 при помощи устройств ввода, таких как клавиатура 40 и указательное устройство 42. Другие устройства ввода (не показаны) могут включать в себя микрофон, джойстик, геймпад, спутниковую антенну, сканер или любое другое.

Эти и другие устройства ввода соединены с процессором 21 часто посредством интерфейса 46 последовательного порта, который связан с системной шиной, но могут быть соединены посредством других интерфейсов, таких как параллельный порт, игровой порт или универсальная последовательная шина (УПШ). Монитор 47 или другой тип устройства визуального отображения также соединен с системной шиной 23 посредством интерфейса, например, видеоадаптера 48. В дополнение к монитору 47, персональные компьютеры обычно включают в себя другие периферийные устройства вывода (не показано), такие как динамики и принтеры.

Компьютер 20 может работать в сетевом окружении посредством логических соединений к одному или нескольким удаленным компьютерам 49. Удаленный компьютер (или компьютеры) 49 может представлять собой другой компьютер, сервер, роутер, сетевой ПК, пиринговое устройство или другой узел единой сети, а также обычно включает в себя большинство или все элементы, описанные выше, в отношении компьютера 20, хотя показано только устройство хранения информации 50. Логические соединения включают в себя локальную сеть (ЛВС) 51 и глобальную компьютерную сеть (ГКС) 52. Такие сетевые окружения обычно распространены в учреждениях, корпоративных компьютерных сетях, Интернете.

Компьютер 20, используемый в сетевом окружении ЛВС, соединяется с локальной сетью 51 посредством сетевого интерфейса или адаптера 53. Компьютер 20, используемый в сетевом окружении ГКС, обычно использует модем 54 или другие средства для установления связи с глобальной компьютерной сетью 52, такой как Интернет.

Модем 54, который может быть внутренним или внешним, соединен с системной шиной 23 посредством интерфейса 46 последовательного порта. В сетевом окружении программные модули или их части, описанные применительно к компьютеру 20, могут храниться на удаленном устройстве хранения информации. Надо принять во внимание, что показанные сетевые соединения являются типичными, и для установления коммуникационной связи между компьютерами могут быть использованы другие средства.

В заключение следует отметить, что приведенные в описании сведения являются примерами, которые не ограничивают объем настоящего технического решения, определенного формулой. Специалисту в данной области становится понятным, что могут существовать и другие варианты осуществления настоящего технического решения, согласующиеся с сущностью и объемом настоящего технического решения.

Литература

1. <https://hackernoon.com/attention-mechanism-in-neural-network-30aaf5e39512>.
2. <https://medium.com/@Svnced/a-brief-overview-of-attention-mechanism-13c578ba9129>.
3. "Medical Concept Representation Learning from Electronic Health Records and its

Application on Heart Failure Prediction", Edward Choi, Andy Schuetz, Walter F. Stewart, Jimeng Sun, 11/02/2016.

4. "Graph-based attention model for healthcare representation learning", Edward Choi, Mohammad Taha Bahadori, Le Song, Walter F. Stewart, Jimeng Sun, 2017.

5

(57) Формула изобретения

1. Способ поддержки принятия врачебных решений с использованием математических моделей представления пациентов, выполняемый на сервере, включающий следующие шаги:

- 10 - формируют обучающую выборку, содержащую электронные истории болезни пациентов, сгруппированных по пациенту;
- производят предварительную обработку данных, содержащихся в истории болезней пациентов, выбранных из обучающей выборки;
- преобразовывают обработанные данные в последовательность медицинских фактов по каждому пациенту с использованием медицинских онтологий;
- 15 - производят автоматическую разметку полученной последовательности медицинских фактов по каждому пациенту, используя извлеченные из истории болезни пациента диагнозы или другие интересующие факты;
- производят обучение первичных репрезентаций независимо для каждой из
- 20 модальностей;
- осуществляют обучение совместных репрезентаций;
- производят обучение финальных моделей и параметров агрегации;
- получают историю болезни пациента, не входящую в обучающую выборку;
- производят предварительную обработку данных полученной истории болезни
- 25 пациента;
- преобразовывают предварительно обработанные данные в последовательность медицинских фактов с использованием медицинских онтологий;
- полученный набор фактов отправляют на вход обученным финальным моделям;
- определяют диагноз, а также проводят анализ и прогноз развития заболеваний
- 30 пациента с наибольшей вероятностью, соответствующей предъявленному набору фактов.

2. Способ по п. 1, в котором электронная история болезни пациента включает, по крайней мере, следующие данные: состояние пациента, способы лечения пациента, средства, используемые при лечении пациента, результаты анализов.

35 3. Система поддержки принятия врачебных решений с использованием математических моделей представления пациентов включает, по крайней мере, один процессор, оперативную память, запоминающее устройство, содержащее инструкции, загружаемые в оперативную память и выполняемые, по крайней мере, одним процессором, инструкции содержат следующие шаги:

- 40 - формирование обучающей выборки, содержащей электронные истории болезни пациентов, сгруппированных по пациенту;
- предварительную обработку данных, содержащихся в истории болезней пациентов, выбранных из обучающей выборки;
- преобразование обработанных данных в последовательность медицинских фактов по каждому пациенту с использованием медицинских онтологий;
- 45 - автоматическую разметку полученной последовательности медицинских фактов по каждому пациенту, используя извлеченные из истории болезни пациента диагнозы или другие интересующие факты;

- обучение первичных репрезентаций независимо для каждой из модальностей;
- обучение совместных репрезентаций;
- обучение финальных моделей и параметров агрегации;
- получение истории болезни пациента, не входящей в обучающую выборку;
- 5 - предварительную обработку данных полученной истории болезни пациента;
- преобразование предварительно обработанных данных в последовательность медицинских фактов с использованием медицинских онтологий;
- отправку полученного набора фактов на вход обученным финальным моделям;
- определение диагноза, а также анализа и прогноза развития заболеваний пациента
- 10 с наибольшей вероятностью, соответствующей предъявленному набору фактов.

15

20

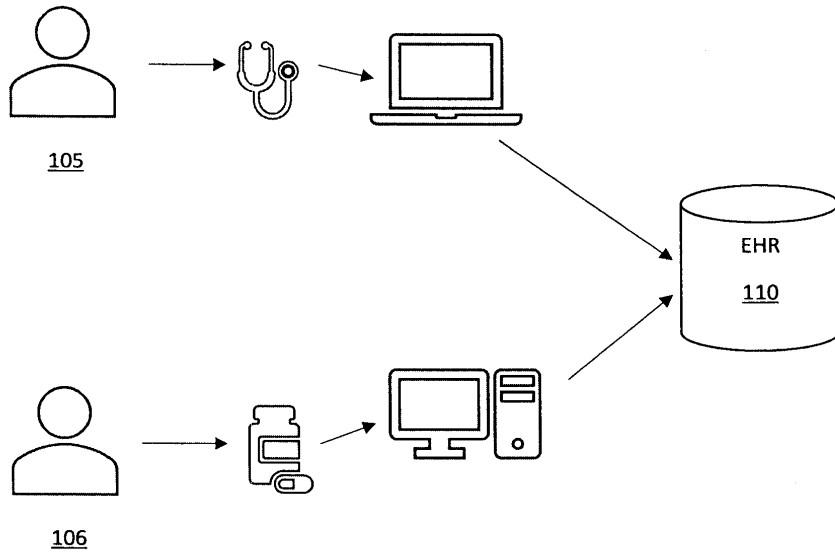
25

30

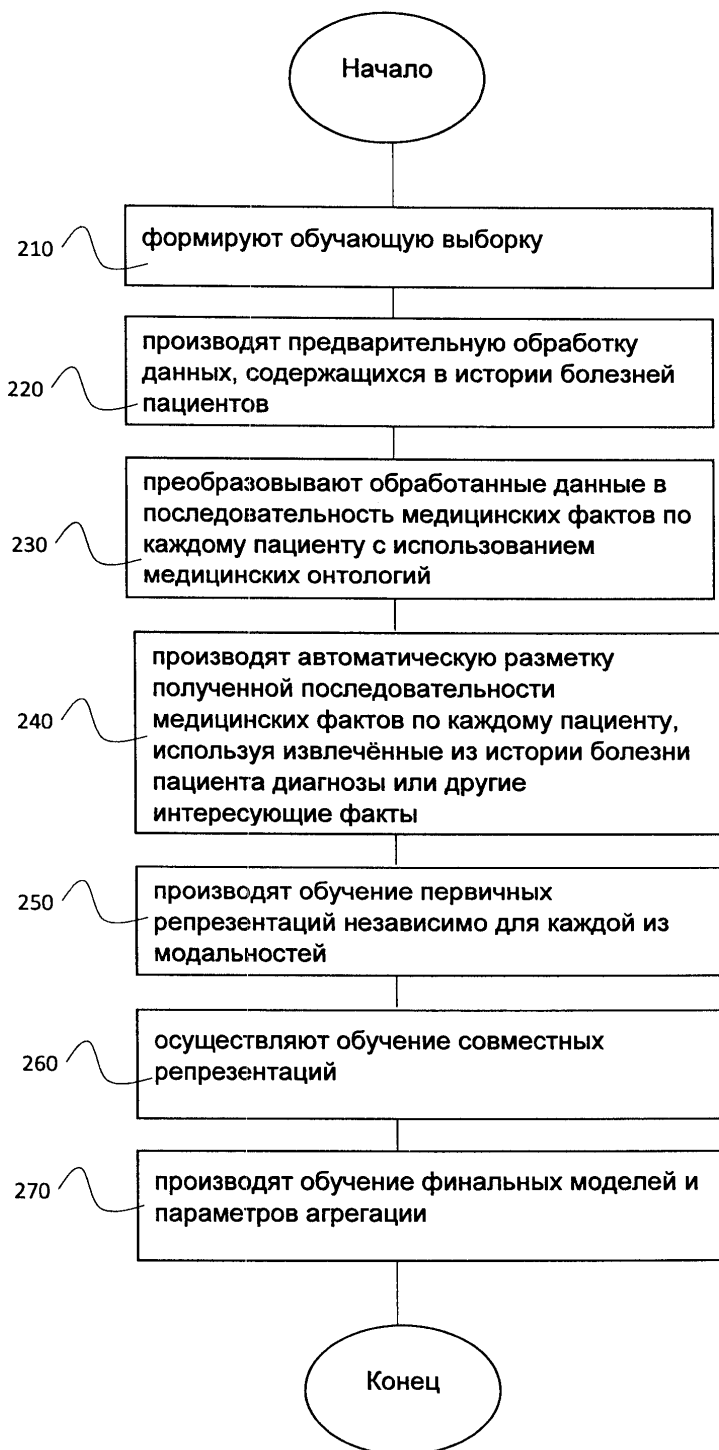
35

40

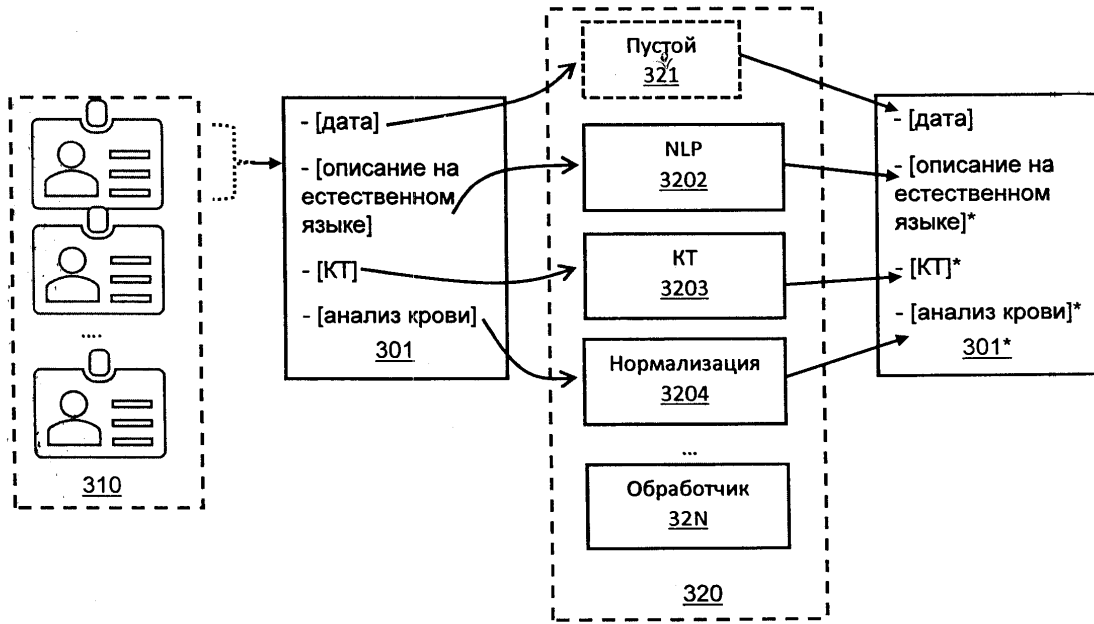
45



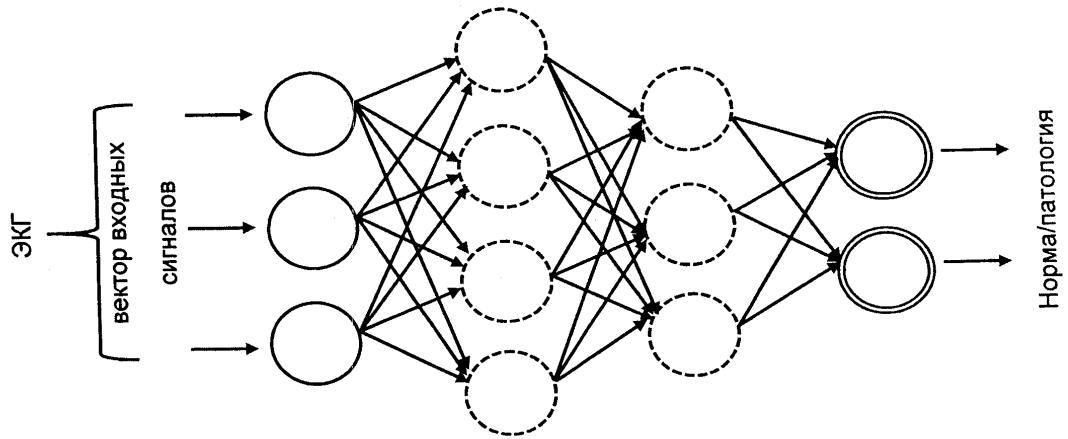
ФИГ. 1



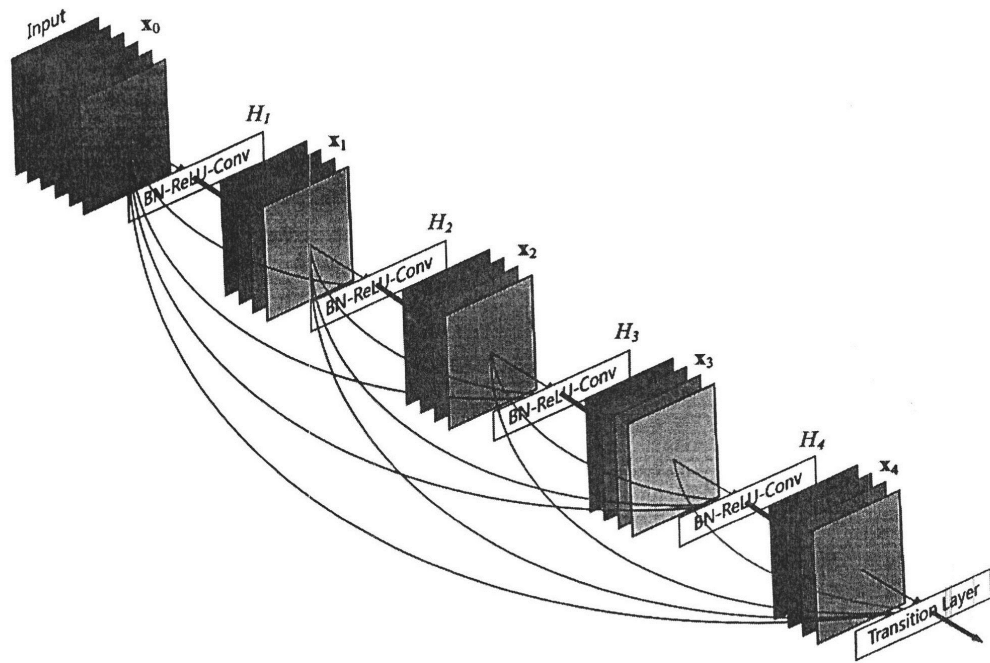
Фиг.2



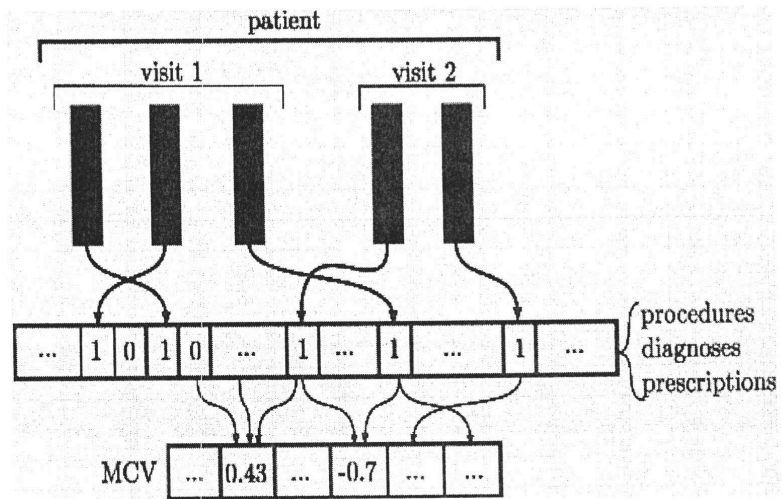
Фиг.3



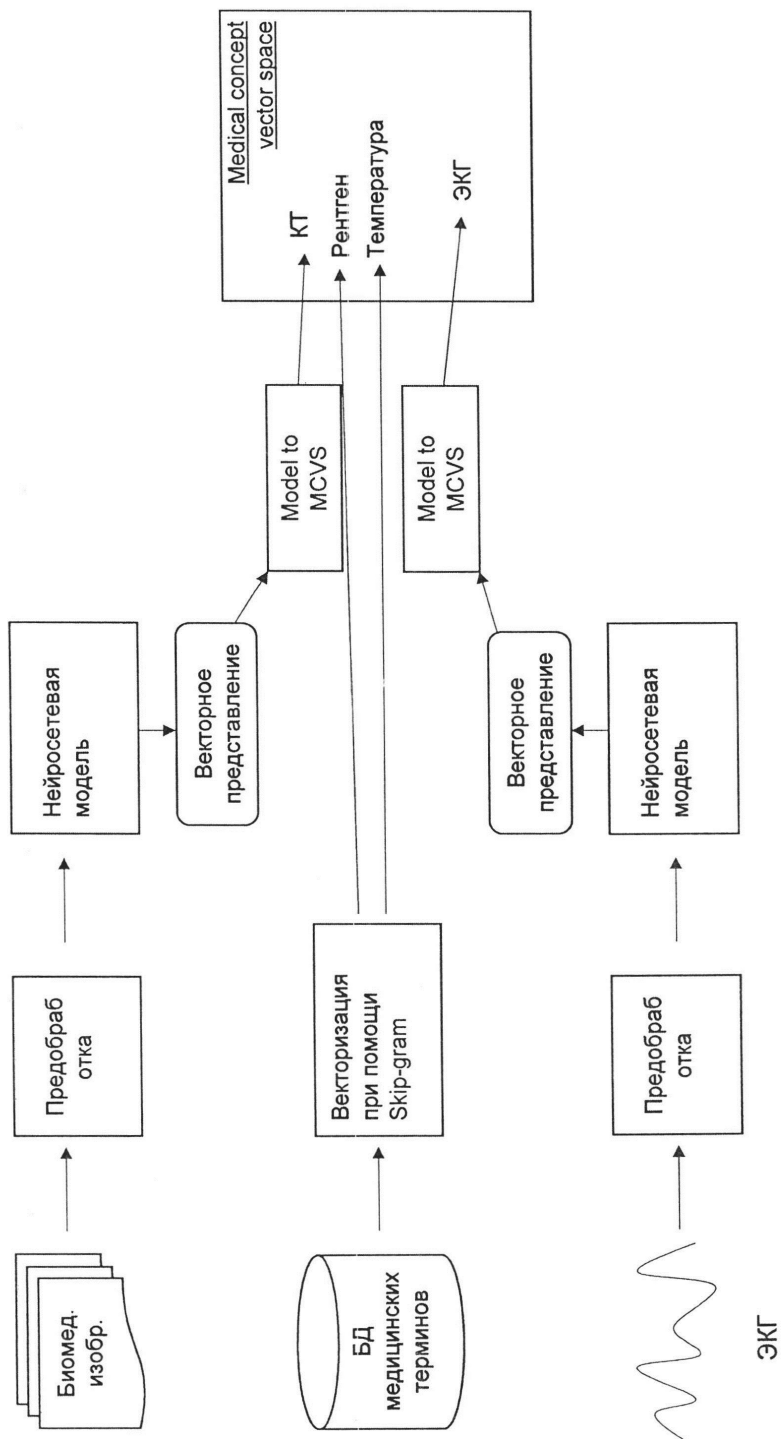
Фиг.4



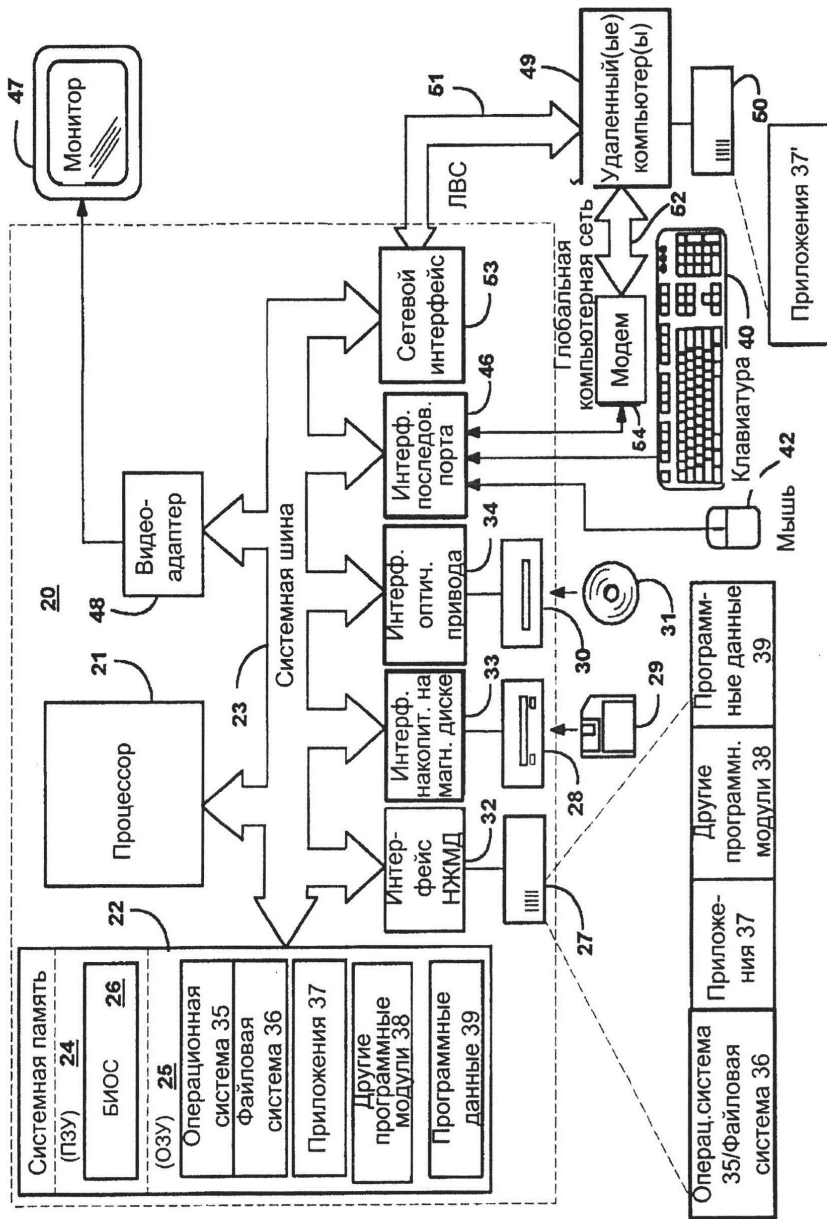
Фиг. 5



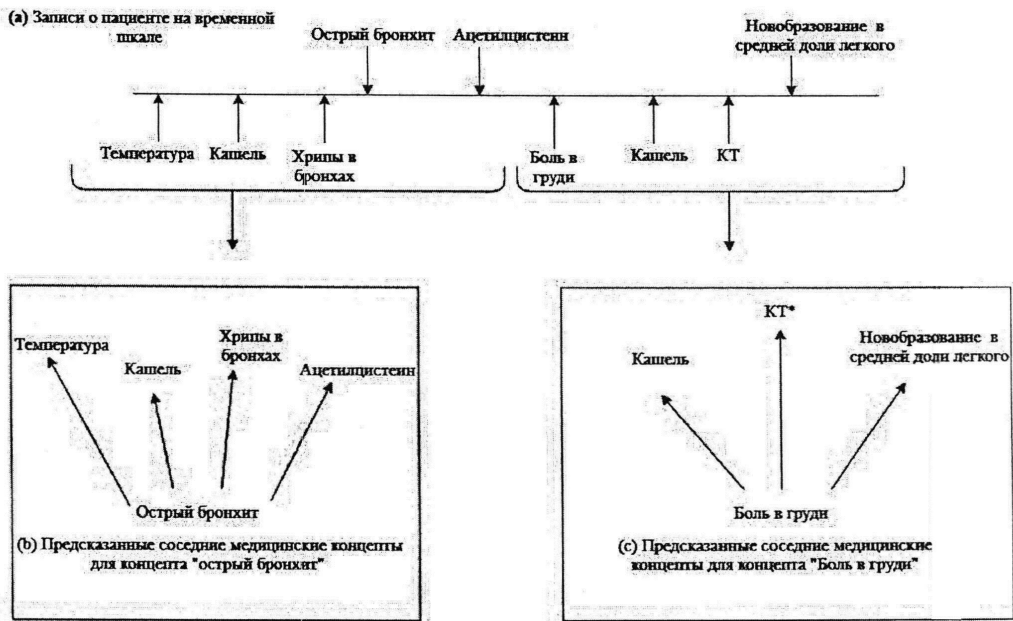
Фиг. 6



Фиг. 7

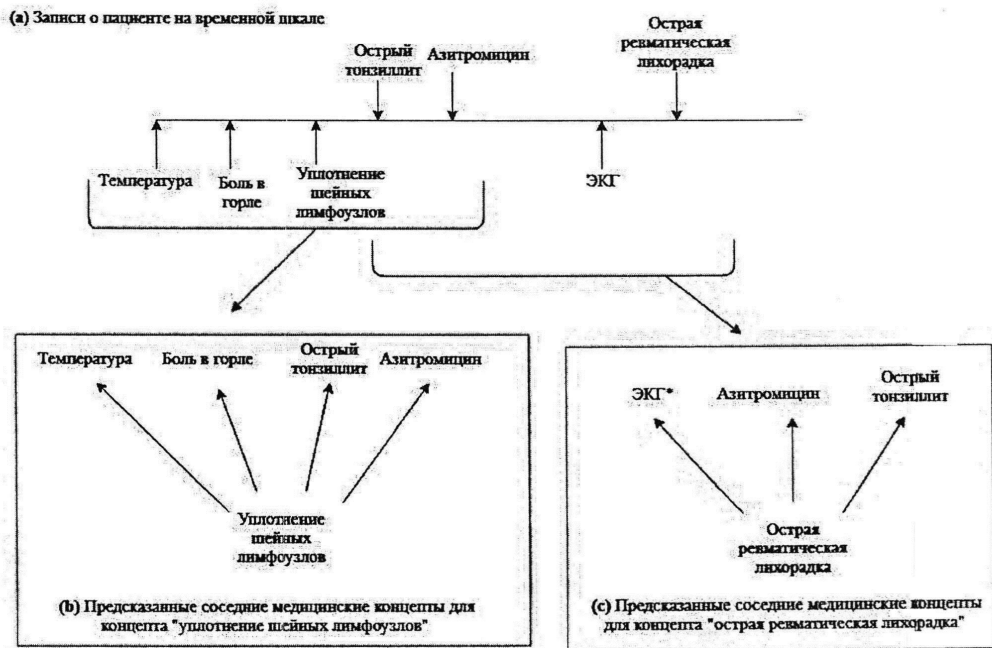


Фиг. 8



*КТ - представляет собой вектор, порожденный нейронной сетью из нативного изображения

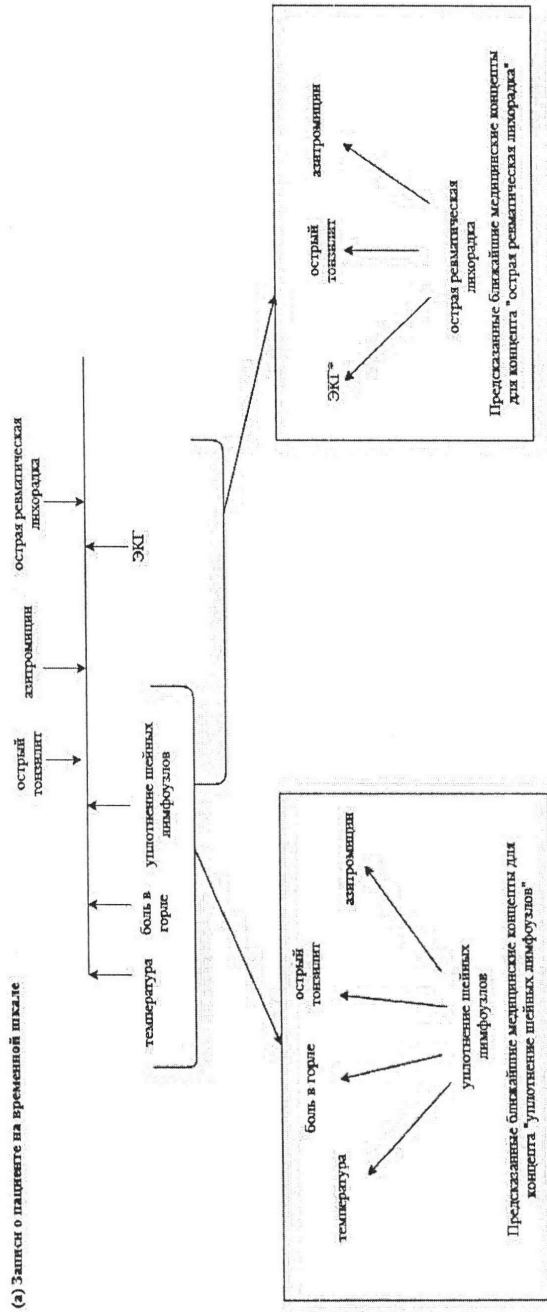
Фиг. 9а



*ЭКГ - представляет собой вектор, порожденный нейронной сетью из нативного сигнала

Фиг. 9б

Обучение совместных репрезентаций



* ЭКГ - представляется как вектор полученный из названного сигнала обработанный нейронной сетью из первичных репрезентаций

(б) Обучение нейронной сети

