

(19)



**Евразийское
патентное
ведомство**

(11) **044383**

(13) **B1**

(12) **ОПИСАНИЕ ИЗОБРЕТЕНИЯ К ЕВРАЗИЙСКОМУ ПАТЕНТУ**

(45) Дата публикации и выдачи патента
2023.08.23

(21) Номер заявки
202390214

(22) Дата подачи заявки
2022.12.30

(51) Int. Cl. **G06F 16/29** (2019.01)
G06F 17/00 (2019.01)
G06Q 40/00 (2023.01)

(54) **СПОСОБ ДЛЯ ОПРЕДЕЛЕНИЯ РАЗМЕЩЕНИЯ ОБЪЕКТОВ ИНФРАСТРУКТУРЫ
БАНКА**

(43) **2023.08.22**

(96) **2022000150 (RU) 2022.12.30**

(71)(73) Заявитель и патентовладелец:
**БАНК ВТБ (ПУБЛИЧНОЕ
АКЦИОНЕРНОЕ ОБЩЕСТВО) (RU)**

(56) WO-A1-2010071477
WO-A1-2019125201
RU-C2-2222046
WO-A1-2017058905
WO-A1-2019164727

(72) Изобретатель:
**Шамаев Юлий Анатольевич, Суржко
Денис Андреевич, Кулик Вадим
Валерьевич, Коновалихин Максим
Юрьевич (RU)**

(74) Представитель:
Котлов Д.В. (RU)

(57) Изобретение относится к способу определения размещения объектов инфраструктуры с использованием алгоритмов кластеризации графов и машинного обучения. Технический результат предлагаемого решения заключается в повышении автоматизации процесса выбора локации для размещения, закрытия или переноса объектов инфраструктуры банка. Предлагаемый способ включает этапы получения и обработки геоинформационных данных, включающих преобразование текстовой информации местоположения в координаты, отображение полученных координат на полученных изображениях заданной территории и расчёт расстояния и времени в пути между отмеченными координатами на полученных изображениях заданной территории и выбор параметра, влияющего на определение размещения объекта инфраструктуры банка. После обработки геоинформационные данные и параметр, влияющий на определение размещения объекта инфраструктуры банка, поступают на вход обученного классификатора для определения объектов инфраструктуры банка. Обучение классификатора включает этап моделирования и кластеризации данных. По результату работы классификатора определяются кластеры, обозначающие благоприятные и неблагоприятные местоположения для размещения объектов инфраструктуры банка, и отображаются на геоинформационных данных.

044383
B1

044383
B1

Область техники

Изобретение относится к области информационных технологий, в частности к способу определения размещения объектов инфраструктуры с использованием алгоритмов кластеризации графов и машинного обучения.

Уровень техники

Управление расположением объектов инфраструктуры, как правило, осуществляется экспертным способом специальным сотрудником. В этой связи была высока вероятность принятия ошибочных решений, и некоторые районы оказывались непокрыты дополнительными офисами/банкоматами, а некоторые наоборот чрезмерно покрыты.

Из источника информации WO 2010/071477 A1, опубликованного 24.06.2010г., известно решение, направленное на автоматизированный отбор объектов недвижимости с использованием индексированной географической информации. Решение позволяет получить технический результат в виде увеличения скорости и надежности обработки данных об объектах недвижимости, расширения функциональных возможностей, повышения наглядности представляемых результатов об объектах недвижимости, а также в осуществлении самой возможности применения многомерной информации об объектах недвижимости, технически ассоциированной с географической картой местности. Этот результат достигается благодаря тому, что автоматизированная система состоит из вычислительного устройства и, по меньшей мере, одного удаленного устройства. При этом удаленное устройство обеспечивает введение в автоматизированную систему критериев отбора объектов недвижимости и преобразование этих критериев в электронный цифровой сигнал. Электронный цифровой сигнал о критериях отбора объектов недвижимости передается удаленным устройством в вычислительное устройство. В соответствии с переданными критериями отбора вычислительное устройство обращается к доступным для него электронным цифровым данным об объектах недвижимости в заданной местности. Среди этих электронных цифровых данных имеются данные об условной пространственной координате, присвоенной каждому объекту недвижимости в заданной местности с использованием принципа индексированного дерева (конечного неориентированного связного графа, не содержащего циклов).

Недостатком решения является наличие удаленного устройства ввода, то есть требует наличие человека, осуществляющего ввод данных об объекте непосредственно в анализируемой локации - таким образом, не решены вопросы высоких трудозатрат анализа и человеческих ошибок ввода данных оператором.

Предлагаемое решение основано на большом массиве статистических данных собираемых из различных внутренних и внешних автоматизированных систем, и источников, таких как данные сотовых операторов, агрегаторов объявлений о продаже недвижимости, операторов фискальных данных и внутренних учетных систем банка, что позволяет достичь высокой репрезентативности, при этом избегая человеческих ошибок.

Из источника информации RU 2679231 C1, опубликованного 06.02.2019г., известно решение геоделирования сети устройств самообслуживания. Технический результат направлен на обеспечение функции автоматизированного определения мест установки устройств самообслуживания (УС) с привязкой к геоинформационным данным. Компьютерно-реализуемый способ геоделирования размещения УС, выполняемый с помощью процессора и содержащий этапы, на которых получают с помощью цифровой карты географическую информацию, определяют на основании упомянутой информации границы населенного пункта, выполняют определение по меньшей мере одной точки места предполагаемой установки по меньшей мере одного УС в упомянутой области, получают набор метрик, содержащий по меньшей мере данные финансовых транзакций в упомянутой области и статистические данные упомянутой области, определяют и ранжируют точки предполагаемых мест установки УС на основании упомянутого набора метрик и отображают определенные одну или более точек предполагаемых мест установки УС на цифровой карте.

Недостатком решения является несистемный подход к определению населения анализируемой области, данные переписи населения, социологические исследования, камеры наблюдения с функцией видеоаналитики могут не обладать достаточной точностью и устойчивостью к допущениям ошибок.

В предлагаемом решении в качестве источника данных о населении используются данные сотовых операторов, позволяющие оценить не только плотность населения анализируемой области, но и его социально-экономические показатели, такие как доходы, при этом обеспечена высокая точность анализа в области не более 500×500 метров.

Сущность изобретения

Технической задачей, на решение которой направлено заявленное техническое решение, является создание способа для определения размещения объектов инфраструктуры банка с использованием алгоритмов кластеризации графов и машинного обучения, охарактеризованного в независимом пункте формулы изобретения, которые способствуют более точному определению местоположения для размещения объектов инфраструктуры банка. Дополнительные варианты реализации настоящего изобретения представлены в зависимых пунктах формулы изобретения.

Техническим результатом, достигающимся при решении вышеуказанной технической задачи, явля-

ется повышение автоматизации процесса выбора локации для размещения, закрытия или переноса объектов инфраструктуры банка. Дополнительным техническим результатом является реализация назначения изобретения, а именно в определении размещения объектов инфраструктуры банка. Дополнительным техническим результатом является сокращение времени обработки информации для определения размещения объектов инфраструктуры банка за счет использования алгоритмов машинного обучения.

Заявленный технический результат достигается за счет осуществления способа для определения размещения объектов инфраструктуры банка, выполняющегося на вычислительном устройстве, содержащем процессор и память, хранящую инструкции, исполняемые процессором, и включающего этапы, на которых:

на вычислительном устройстве пользователь вводит текстовую информацию местоположения для определения размещения объектов инфраструктуры банка и направляют ее на сервер, осуществляющий обработку геоинформационных данных,

посредством сервера, осуществляющего обработку геоинформационных данных, осуществляют обработку геоинформационных данных, где получают

изображения поверхности земли, полученные в результате аэро- и космической съемки заданной территории, посредством сервиса геокодирования преобразуют текстовую информацию местоположения для определения размещения объектов инфраструктуры банка в координаты, соответствующие указанному местоположению, посредством сервиса отображения отмечают указанные координаты на полученных изображениях поверхности земли, а также посредством сервиса прокладывания маршрута рассчитывают расстояние и время в пути между отмеченными координатами на изображении для определения близости координат в реальных городских условиях;

обработанные геоинформационные данные передают на сервер, осуществляющий обработку данных для определения разрешения объектов инфраструктуры банка в базу данных;

на сервер, осуществляющий обработку данных для определения размещения объектов инфраструктуры банка, в базу данных поступают данные из внутренних и внешних источников, относящиеся к размещению объектов инфраструктуры, а именно: внешние данные, содержащие по меньшей мере данные, полученные от телекоммуникационных операторов заданной территории, внутренние данные, характеризующие по меньшей мере параметры, влияющие на определение размещения объектов инфраструктуры банка, причем в базе данных полученные данные обрабатываются и преобразуются в географическую сетку, состоящую из набора полигонов;

на сервере, осуществляющем обработку данных для определения размещения объектов инфраструктуры банка, объединяют геоинформационные данные и географическую сетку, состоящую из набора полигонов в единую географическую сетку;

на сервере, осуществляющем обработку данных для определения размещения объектов инфраструктуры банка, посредством модели машинного обучения, осуществляют определение размещения объектов инфраструктуры банка, где на вход обученному классификатору, определяющему размещение объектов инфраструктуры, поступает единая географическая сетка и параметры, влияющие на определение размещения объектов инфраструктуры банка, причём обучение классификатора включает следующие этапы:

используя алгоритм отбора признаков, осуществляют отбор значимых признаков, из единой географической сетки, состоящей из набора полигонов, для построения модели определения размещения объектов инфраструктуры банка,

определяют веса моделей как коэффициенты линейной регрессии, для составления списка локаций для размещения объектов инфраструктуры,

осуществляют интерпретацию работы моделей, в по меньшей мере одной из списка локаций, посредством использования интерпретирующей модели, используя данные из базы данных, а также данные, полученные на предыдущем этапе, и получают модель для предсказания параметров, влияющих на определение размещения объектов инфраструктуры банка в по меньшей мере одном полигоне географической сетки,

определяют модель достижения параметра, влияющего на определение размещения объектов инфраструктуры банка для по меньшей мере одного объекта инфраструктуры на основе данных, полученных на предыдущих этапах, используя алгоритм градиентного бустинга;

на основе единой географической сетки и данных, полученных на предыдущем этапе, строится граф для определения географического расстояния центров локаций и модуля разности рангов, выданных моделью, причём определение географического расстояния центра локаций осуществляется посредством сервиса прокладывания маршрута;

восстанавливают граф из геопространственных данных и оптимизируют параметры метрики схожести узлов в графе, посредством алгоритма оптимизации метрики схожести узлов геопространственного графа, для учёта географической близости и схожести локаций с точки зрения прогнозируемого параметра, влияющего на определение размещения объектов инфраструктуры банка;

осуществляют спектральную кластеризацию полученного на предыдущем этапе геопространственного графа для разделения геопространственного графа на кластеры и осуществляют агрегацию показа-

телей по полученным кластерам;

определяют оптимальное количество кластеров для каждого полигона географической сетки, района, города по спектрограмме графа;

полученные кластеры отображают на геоинформационных данных, причем данные кластеры обозначают благоприятные и неблагоприятные местоположения для размещения объектов инфраструктуры банка.

В частном варианте реализации предлагаемого решения, внешние данные дополнительно включают данные о продаже жилой недвижимости на первичном/вторичном рынке, данные о торговой активности, макроэкономические показатели по городам/регионам.

В другом частном варианте реализации предлагаемого решения, внутренние данные, характеризующие по меньшей мере параметры, влияющие на определение размещения объектов инфраструктуры банка, включают данные о количестве пользователей, которые проживают в местоположении для определения размещения объектов инфраструктуры банка.

Описание чертежей

Реализация изобретения будет описана в дальнейшем в соответствии с прилагаемыми чертежами, которые представлены для пояснения сути изобретения и никоим образом не ограничивают область применения изобретения. К заявке прилагаются следующие чертежи:

фиг. 1 иллюстрирует систему осуществляющую способ для определения размещения объектов инфраструктуры банка;

фиг. 2 иллюстрирует пример работы способа для определения размещения объектов инфраструктуры банка;

фиг. 3 иллюстрирует пример спектрограммы;

фиг. 4 иллюстрирует пример отображения кластеров, характеризующих благоприятные и неблагоприятные местоположения для размещения объектов инфраструктуры банка на геоинформационных данных.

Детальное описание изобретения

В приведенном ниже подробном описании реализации изобретения приведены многочисленные детали реализации, призванные обеспечить отчетливое понимание настоящего изобретения. Однако, квалифицированному в предметной области специалисту, будет очевидно каким образом можно использовать настоящее изобретение, как с данными деталями реализации, так и без них. В других случаях хорошо известные методы, процедуры и компоненты не были описаны подробно, чтобы не затруднять понимание особенностей настоящего изобретения.

Кроме того, из приведенного изложения будет ясно, что изобретение не ограничивается приведенной реализацией. Многочисленные возможные модификации, изменения, вариации и замены, сохраняющие суть и форму настоящего изобретения, будут очевидными для квалифицированных в предметной области специалистов.

Заявленное решение представляет собой способ для определения размещения объектов инфраструктуры банка с использованием алгоритмов кластеризации графов и машинного обучения, который осуществляется посредством системы (Фиг. 1), которая включает следующие компоненты:

по меньшей мере одно вычислительное устройство пользователя, оборудованное средствами ввода/вывода, экраном и процессором;

сервер, осуществляющий обработку геоинформационных данных, выполненный с возможностью получения изображения поверхности земли, полученные в результате аэро- и космической съемки заданной территории и обработки геоинформационных данных, а также передачи информации на сервер обработки данных посредством API, включающий:

сервис геокодирования, который позволяет преобразовать текстовую информацию местоположения (адрес) в координаты, соответствующие указанному местоположению на изображении поверхности земли (карте);

сервис отображения, который позволяет отмечать на изображениях поверхности земли координаты, полученные от сервиса геокодирования;

сервис прокладывания маршрутов, который предназначен для расчета расстояний и времени в пути между отмеченными координатами на изображении (карте) с учетом дорожной сети и пробок, для определения близости координат в реальных городских условиях;

сервер, осуществляющий обработку данных для определения размещения объектов инфраструктуры банка посредством алгоритмов машинного обучения и включающий по меньшей мере одну базу данных, выполненную с возможностью хранения данных, полученных из внутренних и внешних источников, а также обработки и преобразования этих данных в географическую сетку, состоящую из набора полигонов. База данных реализована на базе технологии Hadoop.

Все элементы системы осуществляют прием и передачу информации посредством средств телекоммуникационной связи.

На фиг. 2, на первом этапе (200), на вычислительном устройстве пользователя, вводят текстовую информацию местоположения (адрес) для определения размещения объектов инфраструктуры банка и

направляют ее на сервер, осуществляющий обработку геоинформационных данных.

Сервер, осуществляющий обработку геоинформационных данных, получает изображения поверхности земли (201), полученные в результате аэро- и космической съемки заданной территории или изображения схемы заданной территории (географические карты, полученные от картографических порталов) и текстовую информацию местоположения для определения размещения объектов инфраструктуры банка. Обработка геоинформационных данных включает преобразование текстовой информации местоположения в координаты, отображение полученных координат на полученных изображениях заданной территории и расчёт расстояния и времени в пути между отмеченными координатами на полученных изображениях заданной территории.

Посредством сервиса геокодирования осуществляют прямое геокодирование (202), т.е. преобразуют текстовую информацию местоположения для определения размещения объектов инфраструктуры банка в координаты, соответствующие указанному местоположению. Прямое геокодирование может осуществляться известными из уровня техники алгоритмами.

Посредством сервиса отображения отмечают указанные координаты на полученных изображениях заданной территории (203). После того как были определены координаты их отображают точками на полученных изображениях заданной территории.

Посредством сервиса прокладывания маршрута рассчитывают расстояние и время в пути между отмеченными точками на изображении для определения близости координат в реальных городских условиях (204).

Обработанные геоинформационные данные передают на сервер, осуществляющий обработку данных для определения разрешения объектов инфраструктуры банка в базу данных (205).

На сервере, осуществляющий обработку данных для определения размещения объектов инфраструктуры банка, в базу данных поступают данные из внутренних и внешних источников, относящиеся к размещению объектов инфраструктуры, а именно: внешние данные, содержащие по меньшей мере данные, полученные от телекоммуникационных операторов заданной территории, внутренние данные, характеризующие по меньшей мере параметры, влияющие на определение размещения объектов инфраструктуры банка, причем в базе данных полученные данные обрабатываются и преобразуются в географическую сетку, состоящую из набора полигонов (206).

Данные могут быть предварительно загружены в базу данных или могут поступать из баз данных внешних систем, с которыми сервер может взаимодействовать посредством модуля API.

Внешние данные могут содержать:

данные, полученные от телекоммуникационных операторов заданной территории для определения плотности населения, пола, возрастной группы, интересов по категориям;

количество запусков приложения мобильного банка с привязкой к местности (по GPS-меткам при запуске приложения);

количество клиентов с открытым продуктом, проживающих на местности;

количество клиентов с брокерским (инвестиционным) продуктом, проживающих на местности;

количество уникальных терминалов по транзакциям за месяц на заданной местности;

количество транзакций на снятие наличных за месяц на заданной местности;

среднее количество транзакций на уникального клиента на заданной местности;

среднее количество транзакций на терминал на заданной местности;

данные о продаже жилой недвижимости на первичном/вторичном рынке могут быть предоставлены агрегаторами объявлений;

и прочие обезличенные данные, которые можно сопоставить с географической сеткой.

Внутренние данные характеризуют по меньшей мере параметры, влияющие на определение размещения объектов инфраструктуры банка, выбирают цель, для которой необходимо разместить или наоборот убрать объект инфраструктуры банка, например, данные о количестве пользователей, которые проживают в местоположении для определения размещения объектов инфраструктуры банка, сумма выданных кредитов, количество посетителей в день, полученных с устройств, регулирующих электронную очередь и т.д.

Также для осуществления работы предлагаемого способа определяют параметр, влияющий на определение размещения объектов инфраструктуры банка. Данный параметр может характеризовать, но не ограничиваясь, показатель планируемого посещения отделения банка или банкомата, показатель планируемой реализации продукта банка и т.д.

Для преобразования данных в географическую сетку полученные данные обрабатываются в базе данных. Обработка включает удаление нерелевантных и (или) некорректных значений данных. Для этого в базе данных запускают предварительный расчет признаков, заключающийся в расчёте производных значений, показателей, использующих данные нескольких источников, показателей, агрегированных по географическому/иному признаку. В результате данного расчета, преобразованные данные объединяются в таблицу и преобразуются в географическую сетку, состоящую из набора полигонов, определяемых координатами угловых точек. Таблица может содержать следующие данные, но не ограничиваясь: по-

лигон - наименование населенного пункта, уровень административного деления, уникальный идентификатор полигона, широта левой нижней точки полигона, долгота левой нижней точки полигона, широта верхней левой точки полигона, долгота верхней левой точки полигона, широта правой верхней точки полигона, долгота правой верхней точки полигона, широта правой нижней точки полигона, долгота правой нижней точки полигона.

Данные, представленные в виде точек, например, объявления о продаже недвижимости, необходимо агрегировать по вхождению в соответствующий полигон. Определяют соответствие данных выбранному полигону по данным, содержащимся в самом объявлении, например, но не ограничиваясь:

house_address - адрес объекта недвижимости;
 longitude - долгота;
 latitude - широта;
 area_1 - административный район 1;
 area_2 - административный район 2;
 subarea - административный район 3;
 construction_year - год постройки;
 rooms_number - количество комнат;
 rooms_number_type - типизированное количество комнат;
 room_space - общая площадь;
 living_space - жилая площадь;
 kitchen_space - площадь кухни;
 floor - этаж;
 floors - всего этажей в доме;
 pricejnetet - цена квадратного метра;
 price - общая цена;
 last_action_date - дата последнего действия с объявлением.

Все эти параметры формируют сведения о социально-демографическом уровне того или иного полигона и соответственно влияют на оценку месторасположения объекта инфраструктуры.

В частном варианте реализации, данные из внешних источников могут передаваться уже в виде рассчитанных данных по географической сетке проставленных меток. Например, банк заключает договор на поставку данных с оператором сотовой связи, банк передает оператору географическую сетку для расчета признаков, оператор агрегирует детальную информацию по своим абонентам по сетке банка и предоставляет доступ к API для загрузки результатов в банк. Далее банк получает данные путем обращения к данному сервису, в таком случае данные поступают в алгоритм напрямую и учитываются в географической сетке в виде отдельного слоя, минуя этап геокодирования.

На сервере, осуществляющем обработку данных для определения размещения объектов инфраструктуры банка, объединяют геоинформационные данные (изображения заданной территории с отмеченными точками введенного пользователем адреса), полученные от сервера, осуществляющего обработку геоинформационных данных, и географическую сетку, состоящую из набора полигонов в единую географическую сетку, состоящую из набора полигонов 500×500 метров (207).

На сервере, осуществляющем обработку данных для определения размещения объектов инфраструктуры банка, посредством модели машинного обучения, осуществляют определение размещения объектов инфраструктуры банка, где на вход обученному классификатору, определяющему размещение объектов инфраструктуры, поступает единая географическая сетка и параметры, влияющие на определение размещения объектов инфраструктуры банка (208).

Обучение классификатора включает этап моделирования и этап кластеризации.

Этап моделирования включает обучение моделей машинного обучения на базе алгоритма градиентного бустинга LightGBM, где на вход классификатора поступает единая географическая сетка и параметры, влияющие на определение размещения объектов инфраструктуры банка.

Используя алгоритм отбора признаков, осуществляют отбор значимых признаков, из единой географической сетки, состоящей из набора полигонов, для построения модели определения размещения объектов инфраструктуры банка. Для отбора признаков может использоваться алгоритм "Boruta" или другой известный алгоритм для отбора признаков. При работе данного алгоритма сравнивают важность используемых атрибутов данных с важностью, так называемых "теневых" или случайных атрибутов, получаемых путем копирования и перетасовки с исходными атрибутами. Впоследствии те атрибуты, которые оказывают влияние меньше, чем теневые из расчета убираются.

После алгоритма "Boruta" запускают алгоритм "HSIC", которым решается проблема определения наилучшего подмножества признаков, максимально независимых друг от друга, которые дадут максимальный эффект для дальнейшего расчета.

Комбинация применения данных алгоритмов обусловлена тем, что иногда возможно получить достаточно небольшой объем исходных данных при том, что социально-демографических показателей всегда примерно равное количество, поэтому требуется оценить влияние тех или иных атрибутов на поведение модели. Без этапа отбора признаков возникает риск получить ложный результат расчета.

Таким образом, для моделей отбираются 30-40 признаков (из около 600 исходных признаков) наиболее релевантных параметру, влияющему на определение размещения объектов инфраструктуры банка. Пример определения значимости признака для модели расчета, для параметра, влияющего на определение размещения объектов инфраструктуры банка с большой продажей кредитных продуктов, важнее количество населения с низкими доходами, а для модели продажи депозитных продуктов - с высокими.

После отбора признаков переходят к этапу подбора и оптимизации параметром моделей, для минимизации среднеквадратичной ошибки модели и ее поведения, методом кросс-валидации, то есть перекрестной проверки, когда все имеющиеся в наличии данные разбиваются на N частей, и на N-1 данных производится обучение модели, а оставшаяся часть данных используется для тестирования.

На следующем этапе определяют веса моделей как коэффициенты линейной регрессии, для составления списка локаций для размещения объектов инфраструктуры банка. Веса определяются алгоритмом как коэффициенты линейной регрессии, где, например, в качестве параметра, влияющего на определение размещения объектов инфраструктуры банка, выступает планируемый KPI для того или иного объекта инфраструктуры, например, количество посетителей инфраструктуры банка, а факторами - выполнение плана по каждой из категорий продуктов. Значения коэффициентов затем нормируются (чтобы сумма была равна 1) и округляются до 5%.

Например, пусть Y - это % исполнения общего плана продаж офиса в одной географической местности, X_cred, X_deps, X_comm - это % исполнения плана продаж по соответствующим категориям продуктов, тогда веса w1, w2, w3 линейной регрессии $Y = w1 * X_cred + w2 * X_deps + w3 * X_comm$ определяют веса соответствующих продуктовых моделей в финальном рейтинге локации.

Таким образом, для любого показателя эффективности офиса Y можно подобрать веса w1, w2, w3, соответствующие весам продуктовых моделей в финальном рейтинге локации.

На следующем этапе осуществляют интерпретацию работы моделей в по меньшей мере одной из списка локаций посредством использования интерпретирующей модели, используя данные из базы данных, а также данные полученные на предыдущем этапе, и получают модель для предсказания параметров, влияющих на определение размещения объектов инфраструктуры банка в по меньшей мере одном полигоне географической сетки.

Все оставшиеся данные в базе данных и определенные вводные факторы модели подаются на вход специализированной интерпретирующей модели SHAP (т.е. происходит взаимодействие с базой данных, а также с данными полученными с классификатора), для вычисления вклада каждого из факторов в конечный результат, что позволяет проинтерпретировать результат работы модели в каждой из оцениваемых локаций, а также обосновать перспективность/эффективность локации с точки зрения понятных социально-демографических, финансовых или иных показателей.

Данный этап приводит к интерпретации результатов и таким образом, на входе данного шага есть единая географическая сетка, включающая таблицу, преобразованную в географическую сетку, и параметры, влияющие на определение размещения объектов инфраструктуры банка, для построения модели, а на выходе построенная модель для предсказания параметра, влияющего на определение размещения объектов инфраструктуры банка, в любом полигоне единой географической сетки и интерпретацию данного предсказания в терминах наиболее значимых признаков (геослоев), оказывающих влияние на решение.

На следующем этапе определяют модель достижения параметра, влияющего на определение размещения объектов инфраструктуры банка для по меньшей мере одного объекта инфраструктуры на основе данных, полученных на предыдущих этапах, используя алгоритм градиентного бустинга, основанного на алгоритмах построения дерева решений, в котором устанавливают требуемые параметры и пороги для расчета, например:

colsample_bytree = 0.8 - доля признаков, используемых для построения каждого из решающих деревьев;

importance_type = 'split' - параметр определения важности признака,

определяется как общее количество использований признака при построении деревьев решений;

learning_rate = 0.05 - параметр шага обучения;

max_depth = 9 - максимальная глубина каждого из решающих деревьев;

metric = 'l1' - метрики для подсчета на валидационной выборке, l1 соответствует абсолютному отклонению от истинного значения;

min_child_samples = 20 - минимальное количество значений признака, используемых в вершине каждого из деревьев;

min_child_weight = 0.001 - минимальная сумма весов, используемых в вершине каждого из деревьев;

min_split_gain = 0.0 - минимальное снижение ошибки для дальнейшего ветвления в каждом из деревьев;

n_estimators = 500 - количество базовых решающих деревьев, используемых в алгоритме;

n_jobs = 4 - количество параллельных вычислительных процессов алгоритма;

num_leaves = 300 - максимальное количество листьев в решающих деревьях;

objective = None, целевая функция для алгоритма, используется значение по умолчанию;
 random_state = 777 - случайное состояние для генерации случайных последовательностей, фиксируется для воспроизводимости работы алгоритма;
 reg_alpha = 3 - параметр L1 регуляризации;
 reg_lambda = 5 - параметр L2 регуляризации;
 silent = True - параметр отключает печать сообщений в процессе работы алгоритма;
 subsample = 0.9 - доля обучающей выборки, используемая для построения каждого из решающих деревьев;
 subsample_for_bin= 200000 - количество значений, используемых в процессе бинаризации;
 subsample_freq = 0 - частота взятия обучающих подвыборок, равно 0, не используется.

Представленные введенные параметры приведены в качестве примера и могут быть заменены на другие, таким образом, есть возможность применять описанные алгоритмы при построении прогноза, основываясь на иных входных параметрах и построении иных моделей.

Этап моделирования заканчивается и начинается этап кластеризации, где на основе единой географической сетки и данных, полученных с предыдущего этапа, этапа моделирования, строится граф для определения географического расстояния центров локаций и модуля разности рангов, выданных моделью. Данный граф строится таким образом, что веса ребер рассчитываются обратно пропорционально линейной комбинации географического расстояния центров локаций и модуля разности рангов, выданных моделью. Далее запускают процесс бинаризации для обнуления части весов в соответствии с выбранным предельно допустимым расстоянием (например, 3.5 километра) для связности кластера - что позволяет кастомизировать граф в соответствии с размерами города/области. Определение географического расстояния центра локаций осуществляется посредством сервиса прокладывания маршрута.

Ниже указан пример, как осуществляют развесовку ребер графа.

$$w_{ij} = \frac{1}{1 + D_{ij} + \alpha * T_{ij}}$$

D_{ij} - расстояние между узлами i, j ;

T_{ij} - разность целевой метрики узлов i, j ;

α - свободный параметр.

Для соседних узлов с расстоянием 1 км и значениями целевой метрики 0.9 и 0.8 и коэф. $\alpha=2$, вес ребра будет равен $1/(1+1+2*(0.9-0.8))=0.45$.

На следующем этапе восстанавливают граф из геопространственных данных и оптимизируют параметры метрики схожести узлов в графе, посредством алгоритма оптимизации метрики схожести узлов геопространственного графа, для учёта географической близости и схожести локаций с точки зрения прогнозируемого параметра влияющего на определение размещения объектов инфраструктуры банка.

Данный этап позволяет оптимизировать значение параметра α (из пункта выше - $\alpha=2$), чтобы сбалансировать метрику для разных городов.

Например, для города с большей плотностью населения и большим количеством офисов оптимизировать метрику для более географически компактных кластеров, а для городов с меньшей плотностью/большими расстояниями между районами подойдет метрика, объединяющая однородные участки со схожей целевой метрикой.

В качестве параметров метрики выступают коэффициенты линейной комбинации географического расстояния и расстояния в модельной метрике. Оптимизация метрики связей в графе осуществляется в терминах максимизации совокупного ранга существующей офисной сети, таким образом, что текущие точки продаж имеют наибольший ранг в полученной кластеризации. Ранг точки рассчитывается как сумма рангов, кластеров с коэффициентами обратно пропорциональными расстоянию от офиса до кластера.

На следующем этапе осуществляют спектральную кластеризацию полученного на предыдущем этапе геопространственного графа для разделения геопространственного графа на кластеры и осуществляют агрегацию показателей по полученным кластерам. Данный этап реализуется на базе классического алгоритма спектральной кластеризации на сконструированном графе, который позволяет оптимально разделить граф на обоснованное количество кластеров и агрегировать моделируемые и исходные показатели в разрезе полученных кластеров. На данном этапе осуществляется агрегация показателей, что позволяет охарактеризовать локацию в различном масштабе: локально на уровне полигонов 500×500 , на уровне района/кластера и на уровне города/взаимного расположения кластеров для автоматического определения оптимального количества кластеров для каждого полигона, района, города.

На следующем этапе определяют оптимальное количество кластеров для каждого полигона географической сетки, района, города по спектрограмме графа (209).

Например, спектральная кластеризация графа полигонов по г. Новосибирск, позволяет выделить 21 кластер, при отсечении собственных значений матрицы сходства по порогу 0.64 и следующей спектрограммой (Фиг. 3):

По оси X отложено количество предполагаемых кластеров в графе; по оси Y отранжированы собственные значения матрицы сходства графа.

Таким образом, при отсечении собственных значений по оси Y по границе 0.64 граф распадается на 21 кластер (количество собственных значений по оси X, не превосходящих порог).

Распределение собственных значений матрицы сходства позволяет выбрать границу отсечения, при котором дальнейшее увеличение количества кластеров не будет давать значимого прироста однородности кластеров.

Собственные значения матрицы сходства - это набор чисел, обладающих определенными полезными свойствами (базовый термин из линейной алгебры).

Завершающим этапом работы способа является отображение полученных кластеров, на геоинформационных данных, причем данные кластеры обозначают благоприятные и неблагоприятные местоположения для размещения объектов инфраструктуры банка (210). На фиг. 4 отражены данные до спектральной кластеризации (левая часть) и полученные кластеры в результате осуществления спектральной кластеризации (правая часть). Чем ярче кластер, тем данное местоположение является благоприятным для размещения объектов инфраструктуры банка.

Также по результату работы предлагаемого алгоритма могут формироваться отчеты по эффективности размещения текущей офисной сети банка, где параметрами являются, например:

office_id - уникальный идентификатор точки продаж; office_name - наименование точки продаж;

ИПЭ - интегральный показатель эффективности, рассчитывается аналитиками банка, приведен для информации;

Возраст - количество лет с момента открытия точки продаж; Площадь - общая площадь помещений точки продаж; Клиенты - количество клиентов банка приписанных к точке продаж; Доля в БН - доля клиентов в банкиризованном населении, покрываемом точкой продаж, источник - экспертная оценка аналитиков банка;

Уникальность покрытия - доля покрытия точки продаж, не покрываемая другими точками продаж банка;

Покрытие точки БН - численность банкиризованного населения (возраст 18+, доход более 20 тыс. руб.) покрываемого точкой продаж;

Доля в БН - доля клиентов в банкиризованном населении, покрываемом точкой продаж, источник - геоплатформа;

Эффект закрытия - количество банкиризованного населения, выбывающего из покрытия офисной сетью банка при закрытии точки продаж;

Георейтинг взвешенный - рейтинг офиса, рассчитанный как средний рейтинг кластеров, взвешенный на расстояние от точки продаж до центров кластеров;

Георейтинг кластера - рейтинг кластера соответствующей точки продаж;

Георейтинг локации - рейтинг полигона 500×500м соответствующей точки продаж.

Отчет может составляться в виде таблиц, графиков, в которых может применяться цветная индикация полученных значений показателей, на основе которых пользователю и специалисту будут очевидны результаты расчета об оптимальности расположения того или иного расположения отделения банка. По такому же принципу могут быть сформированы отчеты по эффективности размещения банкоматов, других объектов инфраструктуры организации. Кроме этого, в качестве отчета могут предоставляться сведения о рекомендуемых новых локациях, где выше описанный способ расчета показал наиболее высокие показатели.

С течением времени и по необходимости, перечисленные параметры в отчете по эффективности размещения текущей офисной сети могут меняться.

Таким образом, путем последовательного применения алгоритмов градиентного бустинга в указанных конфигурациях сокращают перечень наиболее релевантных признаков или геослоев для выбранной целевой переменной, таким образом, требуется значительно меньше вычислительной мощности для обучения модели.

Вычислительная система, обеспечивающая обработку данных, необходимую для реализации заявленного решения, в общем случае содержит такие компоненты, как: один или более процессоров, по меньшей мере одну память, средство хранения данных, интерфейсы ввода/вывода, средство ввода, средства сетевого взаимодействия.

При исполнении машиночитаемых команд, содержащихся в оперативной памяти, конфигурируют процессор устройства для выполнения основных вычислительных операций, необходимых для функционирования устройства или функциональности одного, или более его компонентов.

Память, как правило, выполнена в виде ОЗУ, куда загружается необходимая программная логика, обеспечивающая требуемый функционал. При осуществлении работы предлагаемого решения выделяют объем памяти, необходимый для осуществления предлагаемого решения.

Средство хранения данных может выполняться в виде HDD, SSD дисков, рейд массива, сетевого хранилища, флэш-памяти и т.п. Средство позволяет выполнять долгосрочное хранение различного вида информации, например, вышеупомянутых файлов с наборами данных пользователей, базы данных, со-

державших записи измеренных для каждого пользователя временных интервалов, идентификаторов пользователей и т.п.

Интерфейсы представляют собой стандартные средства для подключения и работы периферийных и прочих устройств, например, USB, RS232, RJ45, COM, HDMI, PS/2, Lightning и т.п.

Выбор интерфейсов зависит от конкретного исполнения устройства, которое может представлять собой персональный компьютер, мейнфрейм, серверный кластер, тонкий клиент, смартфон, ноутбук и т.п.

В качестве средств ввода данных в любом воплощении системы, реализующей описываемый способ, может использоваться клавиатура. Аппаратное исполнение клавиатуры может быть любым известным: это может быть, как встроенная клавиатура, используемая на ноутбуке или нетбуке, так и обособленное устройство, подключенное к настольному компьютеру, серверу или иному компьютерному устройству. Подключение при этом может быть как проводным, при котором соединительный кабель клавиатуры подключен к порту PS/2 или USB, расположенному на системном блоке настольного компьютера, так и беспроводным, при котором клавиатура осуществляет обмен данными по каналу беспроводной связи, например, радиоканалу, с базовой станцией, которая, в свою очередь, непосредственно подключена к системному блоку, например, к одному из USB-портов. Помимо клавиатуры, в составе средств ввода данных также может использоваться: джойстик, дисплей (сенсорный дисплей), проектор, тачпад, манипулятор мышь, трекбол, световое перо, динамики, микрофон и т.п.

Средства сетевого взаимодействия выбираются из устройства, обеспечивающего сетевой прием и передачу данных, например, Ethernet карту, WLAN/Wi-Fi модуль, Bluetooth модуль, BLE модуль, NFC модуль, IrDa, RFID модуль, GSM модем и т.п. С помощью средств обеспечивается организация обмена данными по проводному или беспроводному каналу передачи данных, например, WAN, PAN, ЛВС (LAN), Интранет, Интернет, WLAN, WMAN или GSM.

Компоненты устройства сопряжены посредством общей шины передачи данных.

В настоящих материалах заявки было представлено предпочтительное раскрытие осуществления заявленного технического решения, которое не должно использоваться как ограничивающее иные, частные воплощения его реализации, которые не выходят за рамки испрашиваемого объема правовой охраны и являются очевидными для специалистов в соответствующей области техники.

ФОРМУЛА ИЗОБРЕТЕНИЯ

1. Компьютерно-реализуемый способ для определения размещения объектов инфраструктуры банка, выполняющийся на вычислительном устройстве, содержащий процессор и память, хранящую инструкции, исполняемые процессором, и включающие этапы, на которых:

на вычислительном устройстве пользователя вводят текстовую информацию местоположения для определения размещения объектов инфраструктуры банка и направляют ее на сервер, осуществляющий обработку геоинформационных данных,

посредством сервера, осуществляющего обработку геоинформационных данных, осуществляют обработку геоинформационных данных, где получают изображения поверхности земли, полученные в результате аэро- и космической съемки заданной территории, посредством сервиса геокодирования преобразуют текстовую информацию местоположения для определения размещения объектов инфраструктуры банка в координаты, соответствующие указанному местоположению, посредством сервиса отображения отмечают указанные координаты на полученных изображениях поверхности земли, а также посредством сервиса прокладывания маршрута рассчитывают расстояние и время в пути между отмеченными координатами на изображении для определения близости координат в реальных городских условиях;

обработанные геоинформационные данные передают на сервер, осуществляющий обработку данных для определения разрешения объектов инфраструктуры банка в базу данных;

на сервер, осуществляющий обработку данных для определения размещения объектов инфраструктуры банка, в базу данных поступают данные из внутренних и внешних источников, относящиеся к размещению объектов инфраструктуры, а именно: внешние данные, содержащие по меньшей мере данные, полученные от телекоммуникационных операторов заданной территории, внутренние данные, характеризующие по меньшей мере параметры, влияющие на определение размещения объектов инфраструктуры банка, причем в базе данных полученные данные обрабатываются и преобразуются в географическую сетку, состоящую из набора полигонов;

на сервере, осуществляющем обработку данных для определения размещения объектов инфраструктуры банка, объединяют геоинформационные данные и географическую сетку, состоящую из набора полигонов в единую географическую сетку;

на сервере, осуществляющем обработку данных для определения размещения объектов инфраструктуры банка, посредством модели машинного обучения, осуществляют определение размещения объектов инфраструктуры банка, где на вход обученному классификатору, определяющему размещение объектов инфраструктуры, поступает единая географическая сетка и параметры, влияющие на определение размещения объектов инфраструктуры банка, причём обучение классификатора включает следую-

щие этапы:

используя алгоритм отбора признаков, осуществляют отбор значимых признаков, из единой географической сетки, состоящей из набора полигонов, для построения модели определения размещения объектов инфраструктуры банка,

определяют веса моделей как коэффициенты линейной регрессии, для составления списка локаций для размещения объектов инфраструктуры, осуществляют интерпретацию работы моделей, в по меньшей мере одной из списка локаций, посредством использования интерпретирующей модели, используя данные из базы данных, а также данные, полученные на предыдущем этапе, и получают модель для предсказания параметров, влияющих на определение размещения объектов инфраструктуры банка в по меньшей мере одном полигоне географической сетки,

определяют модель достижения параметра, влияющего на определение размещения объектов инфраструктуры банка для по меньшей мере одного объекта инфраструктуры на основе данных, полученных на предыдущих этапах, используя алгоритм градиентного бустинга;

на основе единой географической сетки и данных, полученных на предыдущем этапе, строится граф для определения географического расстояния центров локаций и модуля разности рангов, выданных моделью, причём определение географического расстояния центра локаций осуществляется посредством сервиса прокладывания маршрута;

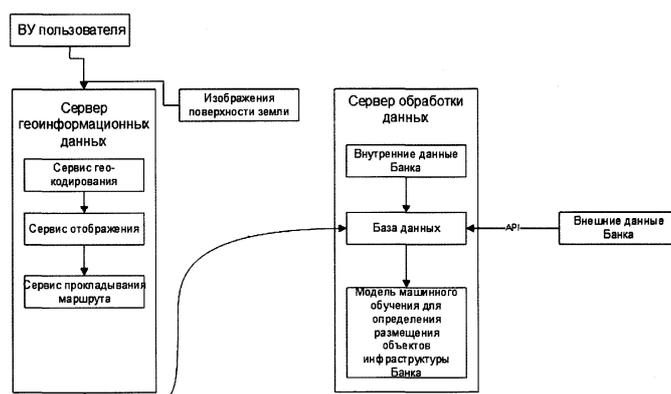
восстанавливают граф из геопространственных данных и оптимизируют параметры метрики схожести узлов в графе, посредством алгоритма оптимизации метрики схожести узлов геопространственного графа, для учёта географической близости и схожести локаций с точки зрения прогнозируемого параметра, влияющего на определение размещения объектов инфраструктуры банка;

осуществляют спектральную кластеризацию полученного на предыдущем этапе геопространственного графа для разделения геопространственного графа на кластеры и осуществляют агрегацию показателей по полученным кластерам;

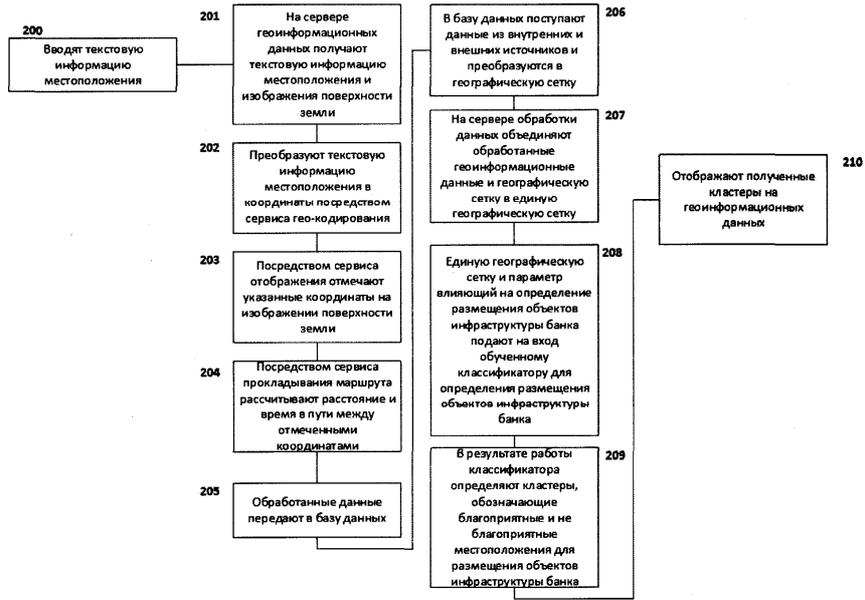
определяют оптимальное количество кластеров для каждого полигона географической сетки, района, города по спектрограмме графа; полученные кластеры отображают на геоинформационных данных, причём данные кластеры обозначают благоприятные и неблагоприятные местоположения для размещения объектов инфраструктуры банка.

2. Способ по п.1, отличающийся тем, что внешние данные дополнительно включают данные о продаже жилой недвижимости на первичном/вторичном рынке, данные о торговой активности, макроэкономические показатели по городам/регионам.

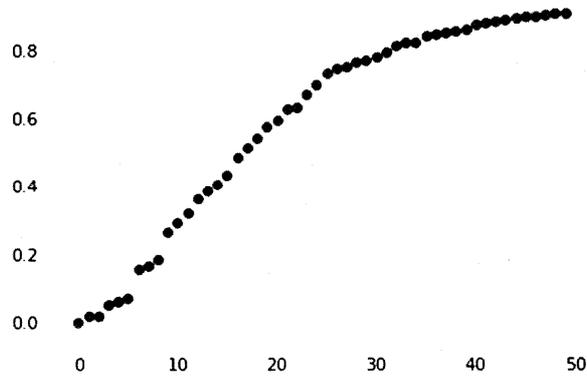
3. Способ по п.1, отличающийся тем, что внутренние данные, характеризующие по меньшей мере параметры, влияющие на определение размещения объектов инфраструктуры банка, включают данные о количестве пользователей, которые проживают в местоположении для определения размещения объектов инфраструктуры банка.



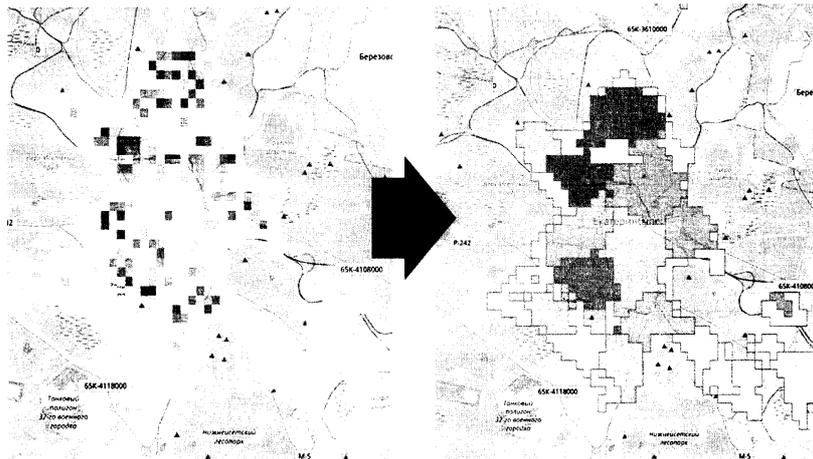
Фиг. 1



Фиг. 2



Фиг. 3



Фиг. 4

