

(19)



**Евразийское
патентное
ведомство**

(21) **201991626** (13) **A1**

(12) ОПИСАНИЕ ИЗОБРЕТЕНИЯ К ЕВРАЗИЙСКОЙ ЗАЯВКЕ

(43) Дата публикации заявки
2020.11.30

(51) Int. Cl. **G06Q 20/40** (2006.01)
G06F 17/00 (2006.01)

(22) Дата подачи заявки
2019.07.31

**(54) СПОСОБ И СИСТЕМА ОПРЕДЕЛЕНИЯ СХОЖЕСТИ ВЕКТОРНЫХ
ПРЕДСТАВЛЕНИЙ УЧАСТНИКОВ ТРАНЗАКЦИЙ**

(31) **2019116394**

(32) **2019.05.28**

(33) **RU**

(71) Заявитель:
**ПУБЛИЧНОЕ АКЦИОНЕРНОЕ
ОБЩЕСТВО "СБЕРБАНК
РОССИИ" (ПАО СБЕРБАНК) (RU)**

(72) Изобретатель:

**Андреев Дмитрий Андреевич, Пинчук
Андрей Михайлович (RU)**

(74) Представитель:

Герасин Б.В. (RU)

(57) Заявленное изобретение относится к области обработки данных, в частности к способу и системе определения схожести векторных представлений участников транзакций. Техническим результатом является обеспечение возможности определения местоположения клиентов на основании данных о транзакции без использования географических координат. Заявленный результат достигается за счет компьютерно-реализуемого способа определения схожести векторных представлений участников транзакций, выполняемого с помощью процессора и содержащего этапы, на которых получают данные транзакций, содержащие, по меньшей мере, идентификационные данные отправителей и получателей транзакций, и данные устройств осуществления транзакций, включающие ID устройств, используемых упомянутыми участниками транзакций; формируют на основании полученных данных цепочки транзакционных активностей, причем каждая из цепочек относится к транзакциям одного участника транзакций между устройствами осуществления транзакций; осуществляют формирование векторных сущностей с помощью преобразования упомянутых цепочек транзакционных активностей, причем упомянутые сущности содержат представления устройств осуществления транзакций; определяют медиану среди значений векторного представления для каждого участника транзакции и формируют его векторную сущность на основании усредненного значения векторных представлений цепочки связанных с ним устройств осуществления транзакции; выполняют расчет косинусного расстояния между векторным представлением участников транзакций, причем участники являются связанными отправителем и получателем транзакции; и определяют векторное расстояние между упомянутыми участниками транзакций на основании значения косинусной близости.

A1

201991626

201991626

A1

СПОСОБ И СИСТЕМА ОПРЕДЕЛЕНИЯ СХОЖЕСТИ ВЕКТОРНЫХ ПРЕДСТАВЛЕНИЙ УЧАСТНИКОВ ТРАНЗАКЦИЙ

ОБЛАСТЬ ТЕХНИКИ

[0001] Заявленное решение относится, в общем, к области обработки данных, а в частности к способу и системе определения схожести векторных представлений участников транзакций.

УРОВЕНЬ ТЕХНИКИ

[0002] В настоящее время в наибольшем количестве случаев мошенничества в качестве канала вывода средств используются переводы на платежные карты (дебетовые или кредитные). Для эффективного противодействия мошенничеству, в частности с учетом доминирующего типа мошенничества «самопереводы», важной является задача определения пространственной геопозиционной близости клиентов, совершающих транзакционные переводы, т.е. анализ их фактического расположения в той или иной географической области.

[0003] Поскольку в большинстве случаев мошенники не знают территориальной принадлежности клиента и не могут подобрать мошеннические карты вывода из той же локации, то осуществляют вывод денежных средств на доступные «дроперские» платежные карты. При этом легитимные переводы в основном происходят между клиентами из близкой геолокации (покупка товаров, оплата оказанных р2р услуг, перевод знакомым и пр.). Информация, позволяющая определить насколько отправитель и получатель близки относительно их геопозиций, поможет повысить качество моделей выявления мошенничества.

[0004] Известные подходы основываются, например, на привязке клиентов к тому или иному территориально расположенному банку и/или геоданные POS-терминалов/УС (банкоматов), в которых клиент пользуется платежной картой, и использовать их для определения геолокации клиента. Такие решения, например, раскрываются в следующих патентных документах: US 20120215701 A1 (Playspan Inc., 23.08.2012), US 20190043054 A1 (Capital One Services LLC, 07.02.2019), US 20120209773 A1 (PayPal Inc., 16.08.2012).

[0005] Известный подход обладает следующими недостатками:\

- Территориальная принадлежность банка – очень обширная территория, поэтому сильно возрастает шанс, что карта вывода мошенников окажется из того же территориального банка;
- POS – геоданные по POS терминалам отсутствуют, есть только информация по мерчанту-арендатору и его юридический адрес/индекс. Данная информация для крупных сетей не позволяет установить местонахождение POS-терминалов, также реально они могут самим мерчантом перемещаться в другие города/регионы без информирования об этом банка;
- Банкомат - по данной категории информация представлена наиболее полно. Проблема заключается в том, что не все клиенты при осуществлении транзакций пользуются банкоматами.

[0006] Таким образом, необходимо разработать эффективный механизм определения геопозиционной близости участников транзакций без использования географических координат.

РАСКРЫТИЕ ИЗОБРЕТЕНИЯ

[0007] Решаемой технической проблемой или технической задачей является определение геопозиционной близости между участниками транзакции на основании их векторных представлений.

[0008] Техническим результатом является обеспечение возможности определения местоположения клиентов на основании данных о транзакции без использования географических координат.

[0009] Основной задачей заявленного способа является представление участников транзакции в виде векторов, позволяющих определить близость/удаленность участников (отправителей и получателей платежей) за счет преобразования их данных в векторную форму и определению векторной близости (например, cosine similarity) и использование данной информации в моделях оценки риска транзакций.

[0010] Заявленный результат достигается за счет компьютерно-реализуемого способа определения схожести векторных представлений участников транзакций, выполняемый с помощью процессора и содержащий этапы, на которых:

– получают данные транзакций, содержащие по меньшей мере идентификационные данные отправителей и получателей транзакций, и данные устройств осуществления транзакций, включающие ID устройств, используемых упомянутыми участниками транзакций;

- формируют на основании полученных данных цепочки транзакционных активностей, причем каждая из цепочек относится к транзакциям одного участника транзакций между устройствами осуществления транзакций;
- осуществляют формирование векторных сущностей с помощью преобразования упомянутых цепочек транзакционных активностей, причем упомянутые сущности содержат представления устройств осуществления транзакций;
- определяют медиану среди значений векторного представления для каждого участника транзакции и формируют его векторную сущность на основании усредненного значения векторных представлений цепочки связанных с ним устройств осуществления транзакции;
- выполняют расчет косинусного расстояния между векторным представлением участников транзакций, причем участники являются связанными отправителем и получателем транзакции; и
- определяют векторное расстояние между упомянутыми участниками транзакций на основании значения косинусной близости.

[0011] В одном из частных вариантов осуществления способа устройства осуществления транзакций представляют собой банкомат и/или POS-терминал.

[0012] В другом частном варианте осуществления способа данные транзакций характеризуют р2р переводы.

[0013] Заявленный результат также реализуется за счет системы определения схожести векторных представлений участников транзакций, которая содержит по меньшей мере один процессор и память, хранящую машиночитаемые инструкции, которые при их исполнении процессором реализуют вышеуказанный способ.

КРАТКОЕ ОПИСАНИЕ ЧЕРТЕЖЕЙ

[0014] Фиг. 1 иллюстрирует блок-схемы процесса выполнения заявленного способа.

[0015] Фиг. 2 иллюстрирует пример валидации тестовой выборки.

[0016] Фиг. 3 иллюстрирует график распределения типов транзакций.

[0017] Фиг. 4 иллюстрирует пример вычислительной системы.

ОСУЩЕСТВЛЕНИЕ ИЗОБРЕТЕНИЯ

[0018] На Фиг. 1 представлен процесс исполнения заявленного способа **(100)** определения схожести векторных представлений участников транзакций. Под

участниками транзакций понимаются лица, осуществляющие транзакционные переводы типа «клиент-клиент» (p2p).

[0019] На первом этапе (101) выполнения способа (100) осуществляется сбор транзакционных данных, которые содержат информацию о транзакционной активности клиентов (покупки товаров, переводы, оплата услуг, снятие наличных и пр.). Каждый транзакционный перевод, как правило, характеризуется информацией, идентифицирующей отправителя перевода и ID устройства для осуществления транзакции, в качестве которых может использоваться POS-терминал или банкомат. Дополнительно может учитываться информация об IP-адресах устройств осуществления транзакций.

[0020] Затем на основании полученной информации на этапе (101) формируют цепочки транзакционных активностей, причем каждая из цепочек относится к транзакциям одного участника транзакций между устройствами осуществления транзакций (102). По транзакционной информации для каждого клиента известна последовательность устройств для осуществления транзакций, которые используются для осуществления операций, например:

Клиент 1 - (POS_1, POS_2, POS_3, ATM_5);

Клиент 2 - (ATM_1, POS_10, POES_2, ATM_4).

[0021] Далее с помощью модели машинного обучения осуществляется векторизация ID устройств для каждой цепочки транзакционной активности (103) для формирования векторных представлений ID устройств, которые используются клиентами в ходе выполнения транзакций.

[0022] В данной задаче используется модель семейства векторного преобразования word2vec, в частности Continuous Bag of Words (CBOW), которая широко применяется в задачах NLP. Суть алгоритма заключается в том, что на вход нейронной сети с одним скрытым слоем и выходным слоем подается контекст какого-либо слова, а целевой переменной для оптимизации при этом является само слово. Таким образом, модель учится предсказывать слово по данному контексту.

[0023] Скрытый слой обученной модели используется как эмбединг (англ. Word embedding — (векторное) представление слова), который на практике показал хорошую способность обобщать взаимоотношения между словами корпуса. В данном случае в качестве «слов» выступают векторные представления устройств для выполнения транзакций: POS/УС (устройства самообслуживания), а «предложений» - последовательность устройств, которые использует один пользователь. Под УС в данном контексте будут пониматься банкоматы.

[0024] Для обучающей выборки модели машинного обучения были выбраны операции в УС и POS-терминалах на подвыборке из пользователей за определенный временной промежуток, в частности 1 месяц. Сэмплирование проводилось по пользователям, соответственно, если пользователь попадает в выборку, то все его УС и POS, которые он использовал для выполнения транзакций, используются для обучения модели. Точки пользователя были упорядочены по времени, и если точка использовалась два или более раз подряд, то повторные использования удалялись, но если эта точка использовалась дальше, после другой точки, то она оставалась в выборке.

Например:

Исходная последовательность точек: A -> A -> C -> A -> B -> B

Последовательность после обработки: A -> C -> A -> B

[0025] Представленный выше пример цепочки используемых ID устройств в векторной форме будет выглядеть следующим образом:

POS_1 = (1,1,1);

POS_2 = (2,0,5);

POS_10 = (1,0,0);

ATM_1 = (3,3,3);

ATM_4 = (1,1,1);

ATM_5 = (0,1,4).

[0026] После преобразования ID устройств для осуществления транзакций в векторную форму, на этапе (104) определяют медиану среди значений векторного представления для каждого участника транзакции и формируют его векторное представление на основании усредненного значения векторных представлений цепочки связанных с ним ID устройств осуществления транзакций.

[0027] Для каждого клиента соответствующая цепочка транзакционной активности приобретает следующий вид: Клиент 1 = ((1,1,1), (2,0,5), (1,1,1), (0,1,4)).

На основании полученного векторного представления в виде цепочки транзакционной активности осуществляется покоординатное усреднение через медиану. Для приведенного выше примера усредненное значение будет представлено в виде вектора (1,1,2.5).

[0028] На основании сформированных цепочек производится трансформация представлений (POS/ATM, IP-адреса) в пространство латентных переменных, в котором косинусная близость усредненных по медиане векторов, описывающих клиентов, определяет их пространственную близость (105), т.е. близость в векторном пространстве между отправителями и получателями транзакции.

[0029] Ниже будет представлен пример вычисления соответствующих векторов.

Пример:

Клиент 1 = (1,1,2.5).

Клиент 2 = (1,1.2,2.1).

Клиент 3 = (4, 0, 0.3)

cosine distance = 1 – cosine similarity

cosine distance Клиент 1 – Клиент 2 = $1 - (1*1 + 1*1.2 + 2.5*2.1)/(2.87*2.62) = 1 - 0.99 = 0.01$

cosine distance Клиент 1 – Клиент 3 = $1 - (1*4 + 1*0 + 2.5*0.3)/(2.87*4.01) = 1 - 0.41 = 0.59$.

[0030] Из приведенного примера видно, что Клиент 1 и Клиент 2 с точки зрения косинусного расстояния расположены гораздо ближе друг к другу по сравнению с Клиентом 1 и Клиентом 3. Следовательно, геопозиционная близость и паттерн поведения Клиента 1 и Клиента 2 достаточно близки, из чего можно судить, что транзакции между ними будут более легитимными по сравнению с транзакциями между Клиентом 1 и Клиентом 3.

[0031] Алгоритм формирования векторных представлений пользователей на основании векторных представлений устройств для выполнения транзакций позволяет вычислить embedding-вектора самого пользователя – участника транзакции, чтобы косинусная близость векторов пользователей отвечала их геолокационной близости. Таким образом, данную информацию можно впоследствии использовать для анализа транзакционной мошеннической активности.

[0032] Далее рассмотрим процесс валидации выборки модели машинного обучения, представленный на Фиг. 2. Для быстрой валидации обученных векторных представлений был использован следующий алгоритм: бралось случайное УС (по которому доступны координаты), и ищались ближайшие к нему соседи УС по векторному представлению (эмбедингу). Точка и соседи визуализировались на карте по их геокоординатам. Аналогично по этим же начальным точкам строились самые удаленные УС. Процесс повторялся для нескольких десятков точек. Такая валидация позволяла наглядно оценить, насколько близко на географической карте лежат друг к другу точки, которые близки на эмбединге.

[0033] Второй вариант проверки — это использование эмбединга по УС/POS для расчета расстояния между пользователями и подсчет статистик по этому расстоянию для анализа ложных срабатываний системы фрод-мониторинга (легитимные транзакции) и

мошеннических операций. Расстояние между пользователями считалось по следующему алгоритму:

- 1) Ко всем устройствам, которые использовал пользователь присоединялись рассчитанные измерения эмбединга (если по какому-то из устройств эмбединг отсутствовал, то он удалялся);
- 2) По измерениям эмбединга бралась медиана для каждого пользователя;
- 3) По медиане эмбединга измерялось косинусное расстояние между пользователями.

[0034] Далее по рассчитанному расстоянию брались перцентили с шагом 5 для операций с маркировкой F (мошенничество), G ложные сработки (легитимные). Для заданного диапазона валидационного периода, например, 1 месяц, была получена следующая статистика, представленная в Таблице1.

Таблица 1

	Резолюция G	Резолюция F
Среднее значение	0.438722	0.605779
10%	0.045853	0.368789
15%	0.078012	0.447429
20%	0.129106	0.500713
25%	0.200715	0.535979
30%	0.279735	0.562158
35%	0.347871	0.584887
40%	0.404844	0.605386
45%	0.451961	0.625186
50%	0.491147	0.640448
55.0%	0.524677	0.655949
60%	0.556372	0.677363
65%	0.586804	0.691357
70%	0.616419	0.710837
75%	0.645766	0.726743
80%	0.675748	0.750359
85%	0.707574	0.770369

	Резолюция G	Резолюция F
90%	0.747735	0.803793
95%	0.803826	0.860480

[0035] Из приведенного выше примера следует, что, если установить значения порога допустимости для легитимных операций, например, расстояние в 0.44, то таким образом будет заблокировано 85% мошеннических операций, но при этом ложные срабатывания можно сократить на более чем 40%, что показывает хорошую разделяющую способность данной метрики.

[0036] Если расстояния по всей выборке разбить на бины и потом проанализировать соотношение мошеннических операций к легитимным (ложные срабатывания), то будет видно, что с увеличением дистанции увеличивается количество мошеннических операций и их доля в бине (Таблица 2). Пример графика представлен на Фиг. 3.

Таблица 2

Перцентиль расстояния	Мошенническая	Легитимная	Соотношение F/G
perc_10	12	166	7%
perc_20	11	170	6%
perc_30	10	170	6%
perc_40	22	157	14%
perc_50	26	154	17%
perc_60	28	151	19%
perc_70	42	138	30%
perc_80	56	123	46%
perc_90	80	100	80%
perc_100	107	73	147%

[0037] В результате реализации заявленного способа (100) был создан эффективный принцип представления и обработки данных для определения геоблизости клиентов без использования геокоординат. Также, использование полученных данных может применяться для целей анализа и противодействия мошенническим операциям.

[0038] На Фиг. 4 представлен пример общего вида вычислительной системы (200) на базе вычислительного устройства (200), которое обеспечивает реализацию заявленного способа или является частью компьютерной системы, например, сервером, обрабатывающим необходимые данные для осуществления способа (100).

[0039] В общем случае, вычислительное устройство (200) содержит объединенные общей шиной информационного обмена один или несколько процессоров (201), средства

памяти, такие как ОЗУ (202) и ПЗУ (203), интерфейсы ввода/вывода (204), устройства ввода/вывода (205), и устройство для сетевого взаимодействия (206).

[0040] Процессор (201) (или несколько процессоров, многоядерный процессор) могут выбираться из ассортимента устройств, широко применяемых в текущее время, например, компаний Intel™, AMD™, Apple™, Samsung Exynos™, MediaTEK™, Qualcomm Snapdragon™ и т.п. Под процессором также необходимо учитывать графический процессор, например, GPU NVIDIA или ATI, который также является пригодным для полного или частичного выполнения способа (100). При этом, средством памяти может выступать доступный объем памяти графической карты или графического процессора.

[0041] ОЗУ (202) представляет собой оперативную память и предназначено для хранения исполняемых процессором (201) машиночитаемых инструкций для выполнения необходимых операций по логической обработке данных. ОЗУ (202), как правило, содержит исполняемые инструкции операционной системы и соответствующих программных компонент (приложения, программные модули и т.п.).

[0042] ПЗУ (203) представляет собой одно или более устройств постоянного хранения данных, например, жесткий диск (HDD), твердотельный накопитель данных (SSD), флэш-память (EEPROM, NAND и т.п.), оптические носители информации (CD-R/RW, DVD-R/RW, BlueRay Disc, MD) и др.

[0043] Для организации работы компонентов устройства (200) и организации работы внешних подключаемых устройств применяются различные виды интерфейсов В/В (204). Выбор соответствующих интерфейсов зависит от конкретного исполнения вычислительного устройства, которые могут представлять собой, не ограничиваясь: PCI, AGP, PS/2, IrDa, FireWire, LPT, COM, SATA, IDE, Lightning, USB (2.0, 3.0, 3.1, micro, mini, type C), TRS/Audio jack (2.5, 3.5, 6.35), HDMI, DVI, VGA, Display Port, RJ45, RS232 и т.п.

[0044] Для обеспечения взаимодействия пользователя с вычислительным устройством (200) применяются различные средства (205) В/В информации, например, клавиатура, дисплей (монитор), сенсорный дисплей, тач-пад, джойстик, манипулятор мышь, световое перо, стилус, сенсорная панель, трекбол, динамики, микрофон, средства дополненной реальности, оптические сенсоры, планшет, световые индикаторы, проектор, камера, средства биометрической идентификации (сканер сетчатки глаза, сканер отпечатков пальцев, модуль распознавания голоса) и т.п.

[0045] Средство сетевого взаимодействия (206) обеспечивает передачу данных устройством (200) посредством внутренней или внешней вычислительной сети, например, Интранет, Интернет, ЛВС и т.п. В качестве одного или более средств (206) может использоваться, но не ограничиваясь: Ethernet карта, GSM модем, GPRS модем, LTE

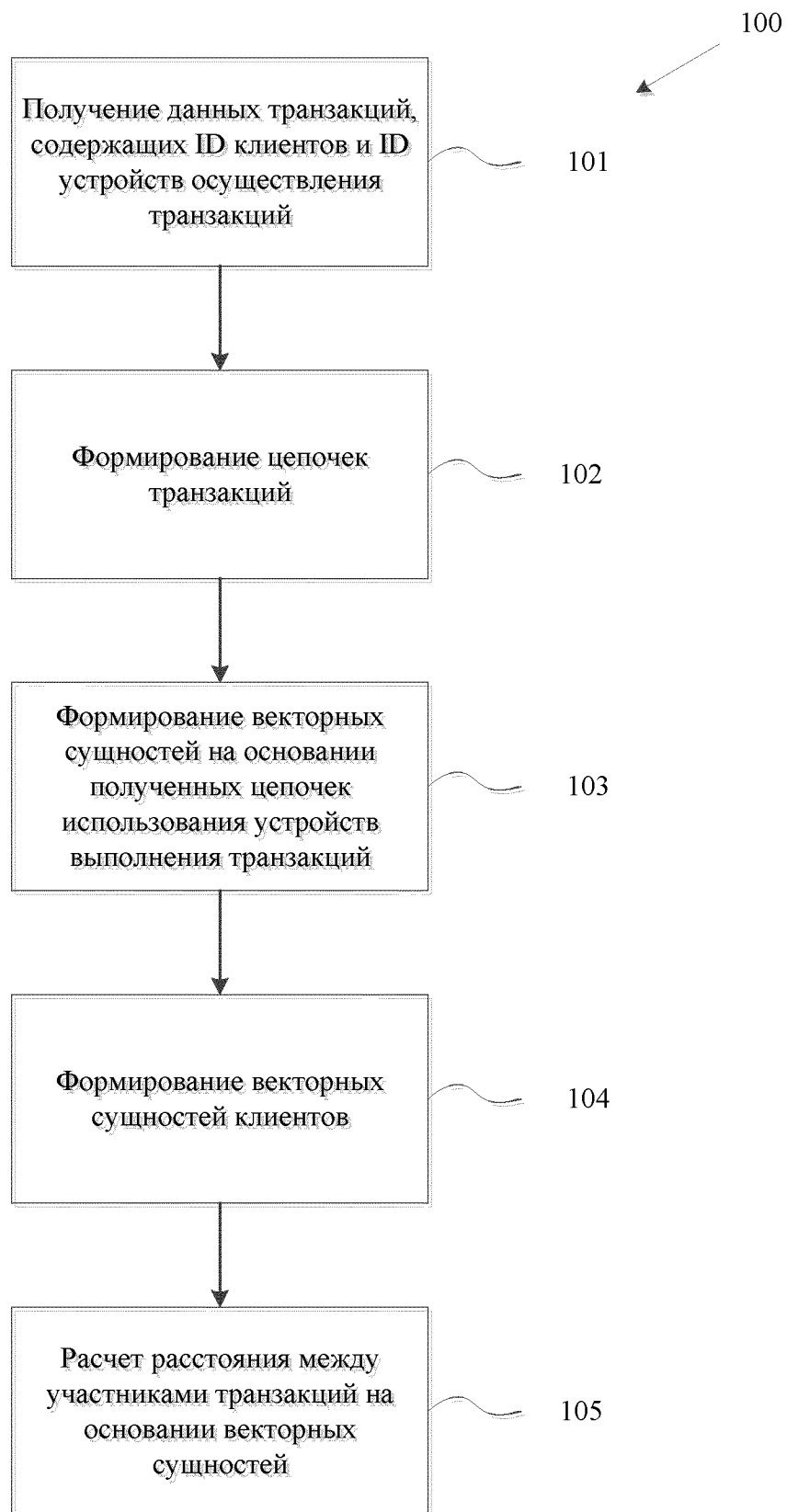
модем, 5G модем, модуль спутниковой связи, NFC модуль, Bluetooth и/или BLE модуль, Wi-Fi модуль и др.

[0046] Дополнительно могут применяться также средства спутниковой навигации в составе устройства (200), например, GPS, ГЛОНАСС, BeiDou, Galileo.

[0047] Представленные материалы заявки раскрывают предпочтительные примеры реализации технического решения и не должны трактоваться как ограничивающие иные, частные примеры его воплощения, не выходящие за пределы испрашиваемой правовой охраны, которые являются очевидными для специалистов соответствующей области техники.

ФОРМУЛА

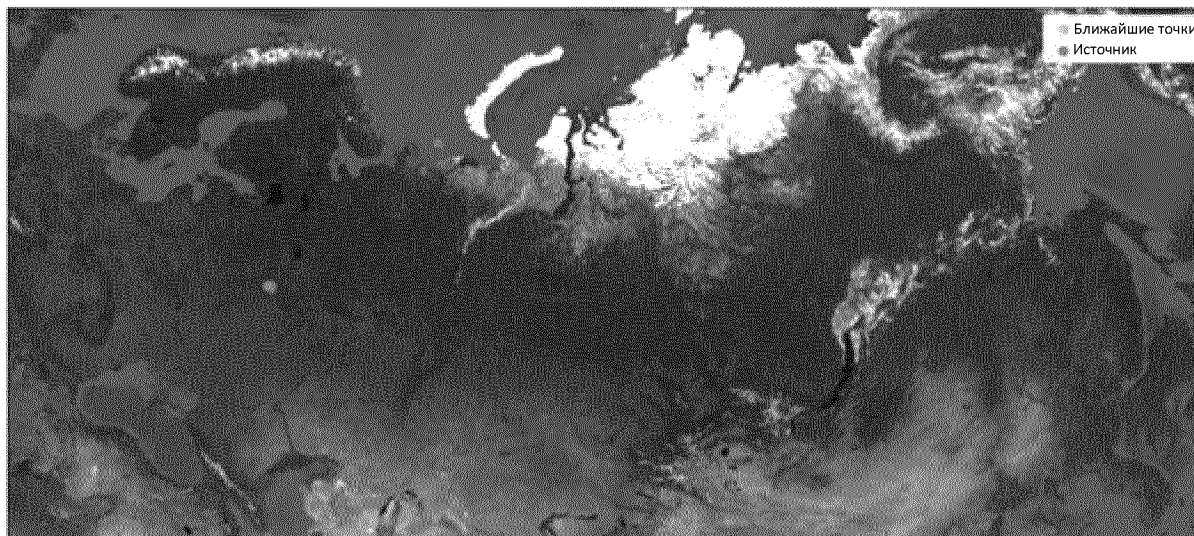
1. Компьютерно-реализуемый способ определения схожести векторных представлений участников транзакций, выполняемый с помощью процессора и содержащий этапы, на которых:
 - получают данные транзакций, содержащие по меньшей мере идентификационные данные отправителей и получателей транзакций, и данные устройств осуществления транзакций, включающие ID устройств, используемых упомянутыми участниками транзакций;
 - формируют на основании полученных данных цепочки транзакционных активностей, причем каждая из цепочек относится к транзакциям одного участника транзакций между устройствами осуществления транзакций;
 - осуществляют формирование векторных представлений устройств осуществления транзакций с помощью преобразования упомянутых цепочек транзакционных активностей;
 - определяют медиану среди полученных значений векторных представлений устройств для осуществления транзакций для каждого участника транзакции и формируют его векторное представление на основании усредненного значения векторных представлений цепочки связанных с ним устройств осуществления транзакции;
 - выполняют расчет косинусного расстояния между усредненными векторными представлениями участников транзакций, причем участники являются связанными отправителем и получателем транзакции; и
 - определяют векторное расстояние между упомянутыми участниками транзакций на основании значения косинусной близости.
2. Способ по п.1, характеризующийся тем, что устройство осуществления транзакций представляет собой банкомат и/или POS-терминал.
3. Способ по п.1, характеризующийся тем, что данные транзакций характеризуют р2р переводы.
4. Система определения схожести векторных представлений участников транзакций, содержащая по меньшей мере один процессор и память, хранящую машиночитаемые инструкции, которые при их исполнении процессором реализуют способ по любому из пп. 1-3.



Фиг. 1

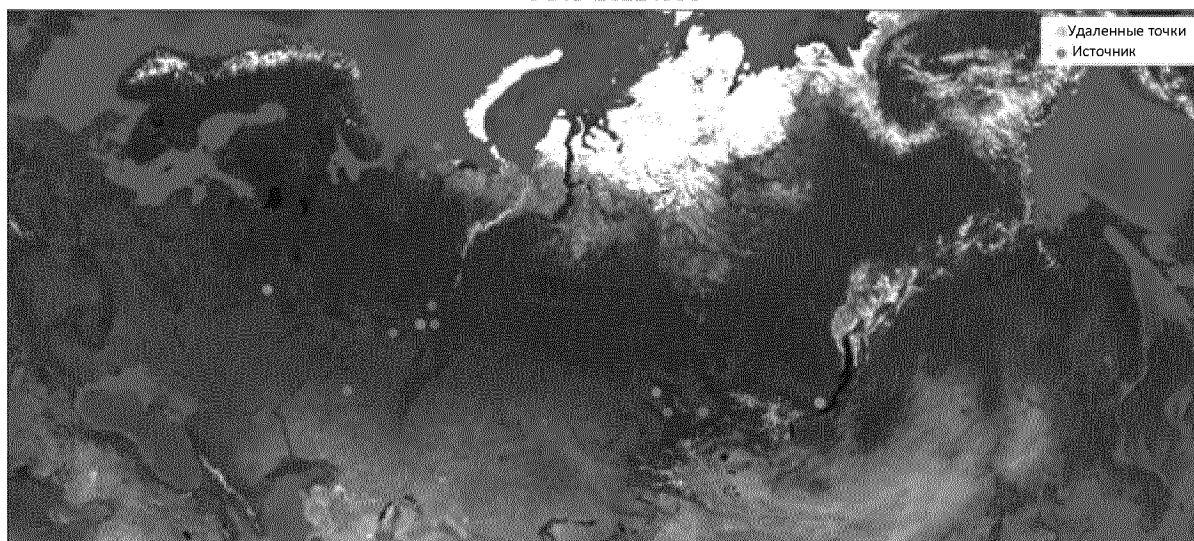
УС ID 26224990

Ближайшие точки

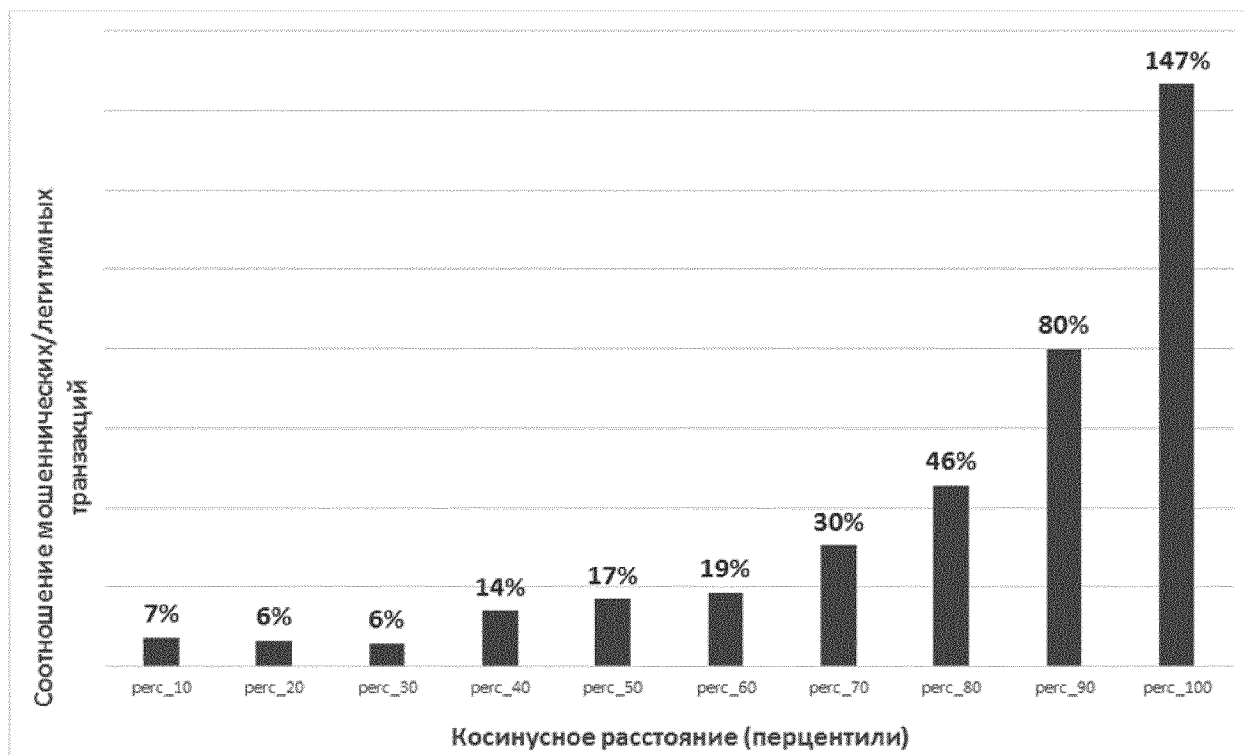


УС ID 26224990

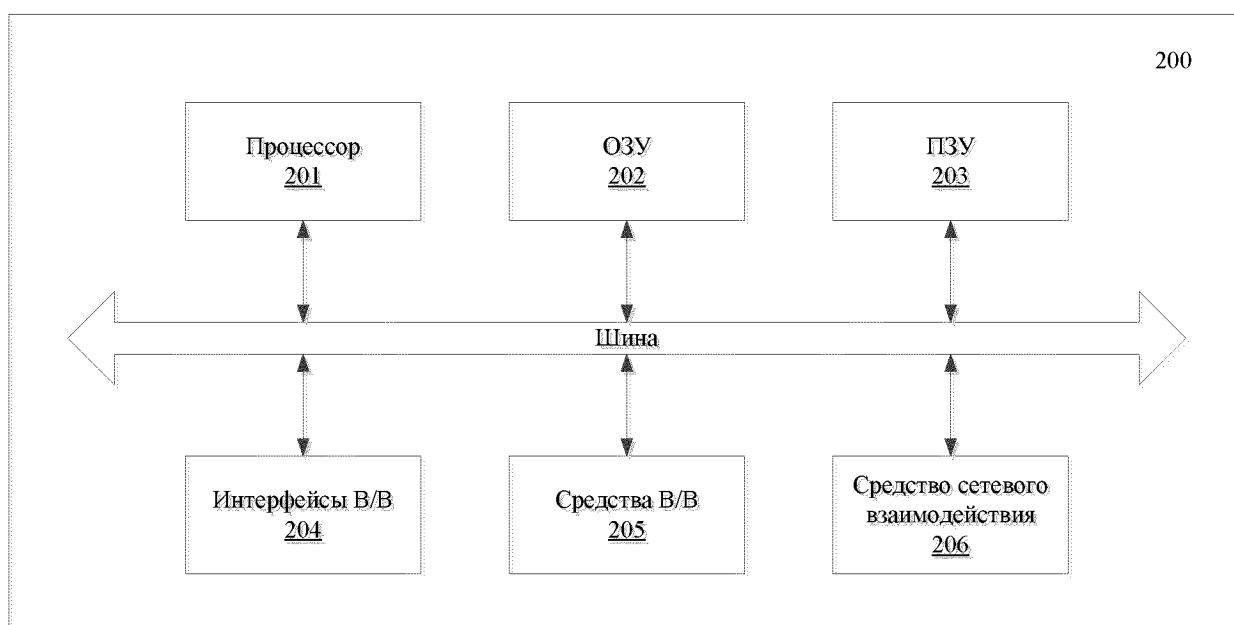
Удаленные точки



Фиг. 2



Фиг. 3



Фиг. 4

ОТЧЕТ О ПАТЕНТНОМ ПОИСКЕ(статья 15(3) ЕАПК
и правило 42 Патентной инструкции к ЕАПК)

Номер евразийской заявки:

201991626**А. КЛАССИФИКАЦИЯ ПРЕДМЕТА ИЗОБРЕТЕНИЯ:****G06Q 20/40 (01/01/2006)****G06F 17/00 (01/01/2006)**

Согласно Международной патентной классификации (МПК)

Б. ОБЛАСТЬ ПОИСКА:

Просмотренная документация (система классификации и индексы МПК)

G06Q 20/00-20/40, G06F 17/00-17/27

Электронная база данных, использовавшаяся при поиске (название базы и, если, возможно, используемые поисковые термины)

В. ДОКУМЕНТЫ, СЧИТАЮЩИЕСЯ РЕЛЕВАНТНЫМИ

Категория*	Ссылки на документы с указанием, где это возможно, релевантных частей	Относится к пункту №
A	US 2017/0169033 A1 (DHRISTI INC) 15.06.2017	1-4
A	RU 2510891 C2 (МАСТЕРКАРД ИНТЕРНЕШНЛ ИНКОРПОРЕЙТЕД) 10.04.2014	1-4
A	RU 2623802 C2 (МАСТЕРКАРД ИНТЕРНЕШНЛ ИНКОРПОРЕЙТЕД) 29.06.2017	1-4
A	US 2017/0193098 A1 (DHRISTI INC) 06.07.2017	1-4

☐ последующие документы указаны в продолжении

* Особые категории ссылочных документов:

"А" документ, определяющий общий уровень техники

"D" документ, приведенный в евразийской заявке

"E" более ранний документ, но опубликованный на дату подачи евразийской заявки или после нее

"O" документ, относящийся к устному раскрытию, экспонированию и т.д.

"P" документ, опубликованный до даты подачи евразийской заявки, но после даты испрашиваемого приоритета"

"Т" более поздний документ, опубликованный после даты приоритета и приведенный для понимания изобретения

"Х" документ, имеющий наиболее близкое отношение к предмету поиска, порочащий новизну или изобретательский уровень, взятый в отдельности

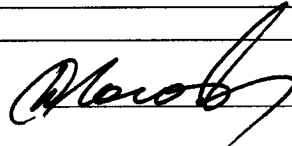
"У" документ, имеющий наиболее близкое отношение к предмету поиска, порочащий изобретательский уровень в сочетании с другими документами той же категории

"&" документ, являющийся патентом-аналогом

Дата проведения патентного поиска: **12/02/2020**

Уполномоченное лицо:

Начальник Управления экспертизы



Д.Ю. Рогожин