

(12) 特許協力条約に基づいて公開された国際出願

(19) 世界知的所有権機関
国際事務局

(43) 国際公開日
2023年3月9日(09.03.2023)

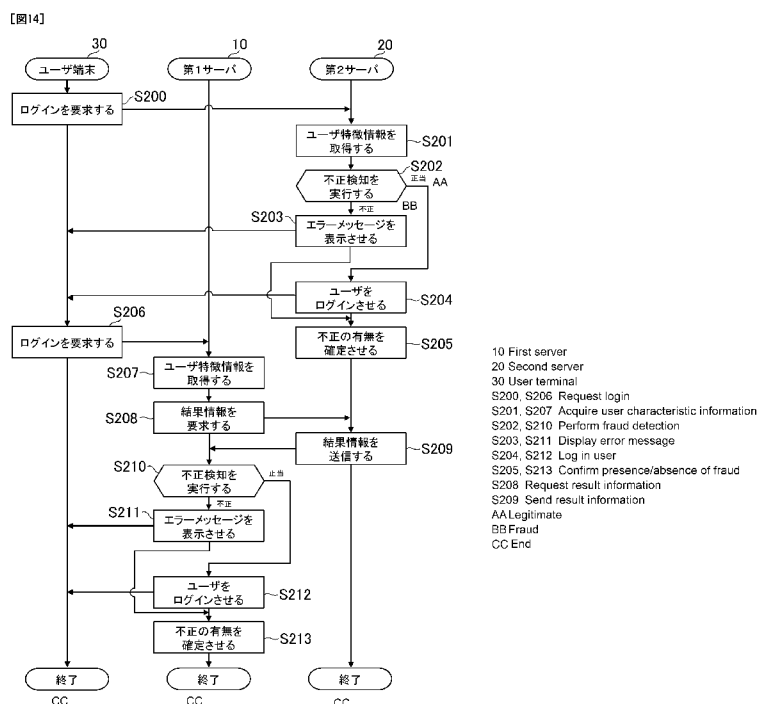


(10) 国際公開番号
WO 2023/032045 A1

- (51) 国際特許分類:
G06F 21/55 (2013.01)
- (21) 国際出願番号: PCT/JP2021/031999
- (22) 国際出願日: 2021年8月31日(31.08.2021)
- (25) 国際出願の言語: 日本語
- (26) 国際公開の言語: 日本語
- (71) 出願人: 楽天グループ株式会社 (RAKUTEN GROUP, INC.) [JP/JP]; 〒1580094 東京都世田谷区玉川一丁目14番1号 Tokyo (JP).
- (72) 発明者: 友田 恭輔 (TOMODA, Kyosuke); 〒1580094 東京都世田谷区玉川一丁目14番1号 楽天グループ株式会社内 Tokyo (JP).
- (74) 代理人: 弁理士法人はるか国際特許事務所 (HARUKA PATENT & TRADEMARK ATTORNEYS); 〒1020085 東京都千代田区六番町3六番町SKビル5階 Tokyo (JP).
- (81) 指定国(表示のない限り、全ての種類の国内保護が可能): AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BN, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CZ, DE, DJ, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IR, IS, IT, JO, JP, KE, KG, KH, KN, KP, KR, KW, KZ, LA, LC, LK, LR, LS, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PA, PE, PG, PH, PL, PT, QA, RO, RS, RU, RW, SA, SC, SD, SE, SG, SK, SL,

(54) Title: FRAUD DETECTION SYSTEM, FRAUD DETECTION METHOD, AND PROGRAM

(54) 発明の名称: 不正検知システム、不正検知方法、及びプログラム



(57) Abstract: A user characteristic information acquisition means (102) of a fraud detection system (S) acquires user characteristic information which relates to characteristics of a user in a first service. A user identification information acquisition means (101) acquires user identification information with which the user can be identified. A result information acquisition means (104) acquires, on the basis of the user identification information, result information that relates to a result of fraud detection for the user in a second service in which a fraud detection engine for detecting fraud differs from

WO 2023/032045 A1

ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG,
US, UZ, VC, VN, WS, ZA, ZM, ZW.

- (84) 指定国(表示のない限り、全ての種類の広域保護が可能): ARIPO (BW, GH, GM, KE, LR, LS, MW, MZ, NA, RW, SD, SL, ST, SZ, TZ, UG, ZM, ZW), ユーラシア (AM, AZ, BY, KG, KZ, RU, TJ, TM), ヨーロッパ (AL, AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, MK, MT, NL, NO, PL, PT, RO, RS, SE, SI, SK, SM, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, KM, ML, MR, NE, SN, TD, TG).

添付公開書類 :

一 国際調査報告 (条約第21条(3))

that of the first service. A fraud detection means (103) detects fraud in the first service on the basis of the user characteristic information in the first service and the result information in the second service.

(57) 要約 : 不正検知システム (S) のユーザ特徴情報取得手段 (102) は、第1サービスにおけるユーザの特徴に関するユーザ特徴情報を取得する。ユーザ識別情報取得手段 (101) は、ユーザを識別可能なユーザ識別情報を取得する。結果情報取得手段 (104) は、ユーザ識別情報に基づいて、不正を検知するための不正検知エンジンが第1サービスとは異なる第2サービスにおけるユーザの不正検知の結果に関する結果情報を取得する。不正検知手段 (103) は、第1サービスにおけるユーザ特徴情報と、第2サービスにおける結果情報と、に基づいて、第1サービスにおける不正を検知する。

明 細 書

発明の名称：不正検知システム、不正検知方法、及びプログラム
技術分野

[0001] 本開示は、不正検知システム、不正検知方法、及びプログラムに関する。

背景技術

[0002] 従来、サービスにおける不正を検知する不正検知システムが知られている。特許文献1には、JSON形式の訓練データに基づいて、教師有り学習の学習モデルの学習を行う技術が記載されている。特許文献2には、学習モデルのデータ構造として、JSON形式のツリー構造を利用する技術が記載されている。特許文献3には、JSON形式のプライバシーポリシーを訓練データとして、学習モデルの学習を行う技術が記載されている。

[0003] 特許文献4には、クラウド上のデータが不正に取得されることを防止するために、JSON形式等の構造化データフォーマットのメタデータを、学習済みの学習モデルに解析させる技術が記載されている。特許文献5には、IoTデバイスの正当性を確保するために、JSON形式等の構造化データで記述された、IoTデバイスからのリクエストの正当性を検証する技術が記載されている。

先行技術文献

特許文献

[0004] 特許文献1：国際公開第2018/139458号公報

特許文献2：特開2021-0811980号公報

特許文献3：特開2020-091814号公報

特許文献4：特開2019-153330号公報

特許文献5：特開2019-009728号公報

発明の概要

発明が解決しようとする課題

[0005] しかしながら、特許文献1-5の技術では、学習モデル又はルールといっ

た不正検知エンジンが個々のサービスで独自に定められているので、他のサービスにおける不正の傾向を活用することができない。あるサービスXにおける不正を検知するために、他のサービスYにおける不正検知の結果を利用することができれば、サービスXにおける不正を検知しやすくなり、セキュリティが高まると考えられる。

[0006] 本開示の目的の1つは、セキュリティを高めることである。

課題を解決するための手段

[0007] 本開示の一態様に係る不正検知システムは、第1サービスにおけるユーザの特徴に関するユーザ特徴情報を取得するユーザ特徴情報取得手段と、前記ユーザを識別可能なユーザ識別情報を取得するユーザ識別情報取得手段と、前記ユーザ識別情報に基づいて、不正を検知するための不正検知エンジンが前記第1サービスとは異なる第2サービスにおける前記ユーザの不正検知の結果に関する結果情報を取得する結果情報取得手段と、前記第1サービスにおける前記ユーザ特徴情報と、前記第2サービスにおける前記結果情報と、に基づいて、前記第1サービスにおける不正を検知する不正検知手段と、を含む。

発明の効果

[0008] 本開示によれば、セキュリティが高まる。

図面の簡単な説明

[0009] [図1]不正検知システムの全体構成の一例を示す図である。

[図2]ユーザが第1サービスを利用する様子の一例を示す図である。

[図3]第1サービスにおける不正検知の流れの一例を示す図である。

[図4]第1実施形態で実現される機能の一例を示す機能ブロック図である。

[図5]第1サービスデータベースの一例を示す図である。

[図6]履歴データベースの一例を示す図である。

[図7]第2サービスデータベースの一例を示す図である。

[図8]不正情報データベースの一例を示す図である。

[図9]第2不正検知エンジンの取得方法の一例を示す図である。

[図10]第1実施形態で実行される処理の一例を示すフロー図である。

[図11]第2実施形態で実現される機能の一例を示す機能ブロック図である。

[図12]第2実施形態における第1サービスの不正検知の一例を示す図である。

。

[図13]履歴データベースの一例を示す図である。

[図14]第1実施形態で実行される処理の一例を示すフロー図である。

[図15]第1実施形態に係る変形例の機能ブロック図である。

[図16]第2実施形態に係る変形例の機能ブロック図である。

発明を実施するための形態

[0010] [1. 第1実施形態]

本開示に係る不正検知システムの実施形態の一例である第1実施形態を説明する。

[0011] [1-1. 不正検知システムの全体構成]

図1は、不正検知システムの全体構成の一例を示す図である。不正検知システムSは、第1サーバ10、第2サーバ20、及びユーザ端末30を含む。ネットワークNは、インターネット又はLAN等の任意のネットワークである。不正検知システムSは、少なくとも1つのコンピュータを含めばよく、図1の例に限られない。

[0012] 第1サーバ10は、第1サービスを提供する第1提供者のサーバコンピュータである。第1サーバ10は、第1サービスにおける不正を検知する。第1サーバ10を含む第1サービスのシステムは、第1不正検知システムといえることができる。制御部11は、少なくとも1つのプロセッサを含む。記憶部12は、RAM等の揮発性メモリと、ハードディスク等の不揮発性メモリと、を含む。通信部13は、有線通信用の通信インタフェースと、無線通信用の通信インタフェースと、の少なくとも一方を含む。

[0013] 第2サーバ20は、第2サービスを提供する第2提供者のサーバコンピュータである。第2サービスは、第1サービスとは異なる他のサービスである。第1実施形態では、第1提供者及び第2提供者が異なる場合を説明するが

、第1提供者及び第2提供者は同じであってもよい。即ち、第2サービスは、第1サービスと同じ提供者が提供する他のサービスであってもよい。なお、提供者は、サービスの運営会社である。

[0014] 第2サーバ20は、第2サービスにおける不正を検知する。第2サーバ20を含む第2サービスのシステムは、第2不正検知システムといえることができる。不正検知システムSは、第1サービスの第1不正検知システムと、第2サービスの第2不正検知システムと、の2つのシステムを含むことになる。制御部21、記憶部22、及び通信部23の物理的構成は、それぞれ制御部11、記憶部12、及び通信部13と同様である。

[0015] ユーザ端末30は、ユーザのコンピュータである。例えば、ユーザ端末30は、パーソナルコンピュータ、スマートフォン、タブレット端末、又はウェアラブル端末である。制御部31、記憶部32、及び通信部33の物理的構成は、それぞれ制御部11、記憶部12、及び通信部13と同様である。操作部34は、タッチパネル等の入力デバイスである。表示部35は、液晶ディスプレイ又は有機ELディスプレイである。

[0016] なお、記憶部12、22、32に記憶されるプログラムは、ネットワークNを介して供給されてもよい。また、各コンピュータには、コンピュータ読み取り可能な情報記憶媒体を読み取る読取部（例えば、メモリカードスロット）と、外部機器とデータの入出力をするための入出力部（例えば、USBポート）と、の少なくとも一方が含まれてもよい。例えば、情報記憶媒体に記憶されたプログラムが、読取部及び入出力部の少なくとも一方を介して供給されてもよい。

[0017] [1-2. 不正検知システムの概要]

第1実施形態では、第1サービスの一例として、画像を投稿することを主な目的としたSNS（Social networking service）を説明する。第2サービスの一例として、短文のメッセージを投稿することを主な目的としたSNSを説明する。第1サービス及び第2サービスは、任意のサービスであってもよく、SNSに限られない。他のサービスの例は、後述の変形例で説明する。

- [0018] 第1実施形態では、ユーザが第1サービス及び第2サービスの両方を利用する場合を説明するが、ユーザは、第1サービス又は第2サービスの何れか一方のみを利用してよい。第1サービスを利用するユーザが少なくとも1人存在し、第2サービスを利用するユーザが少なくとも1人存在すればよい。第1サービスの全てのユーザと、第2サービスの全てのユーザと、が完全に一致する必要はない。
- [0019] 第1実施形態では、ユーザは、第1サービス及び第2サービスに利用登録済みであるものとする。ユーザは、第1サービスにログインするためのユーザID及びパスワードと、第2サービスにログインするためのユーザID及びパスワードと、を発行済みである。例えば、ユーザが、ユーザ端末30のアプリケーション又はブラウザを利用して第1サーバ10にアクセスすると、第1サービスにログインするためのログイン画面が表示部35に表示される。
- [0020] 図2は、ユーザが第1サービスを利用する様子の一例を示す図である。ユーザは、ログイン画面G1の入力フォームF10、F11に、第1サービスのユーザID及びパスワードを入力してボタンB12を選択する。第1サーバ10が、ユーザID及びパスワードの正当性を確認すると、第1サービスへのログインを許可する。第1サービスへのログインが許可されると、第1サービスのホーム画面G2が表示部35に表示される。ユーザは、ホーム画面G2から第1サービスを利用する。ユーザは、第1サービスと同様の流れで第2サービスにログインし、第2サービスを利用できる。
- [0021] 悪意のある者は、フィッシング等によって他人のユーザID及びパスワードを入手し、第1サービス及び第2サービスの少なくとも一方に、他人になりすましてログインすることがある。以降、悪意のある者を不正ユーザと記載する。不正ユーザは、第1サービス及び第2サービスの少なくとも一方で不正をするユーザである。以降、不正をしないユーザを、正当ユーザと記載する。正当ユーザ及び不正ユーザを区別しない時は、単にユーザと記載する。

- [0022] 不正とは、違法行為、利用規約に違反する行為、又はその他の迷惑行為である。第1実施形態では、不正の一例として、なりすまし（不正ログイン）を説明する。このため、なりすましについて説明している箇所は、不正と読み替えることができる。検知対象となる不正自体は、種々の種類であってよく、なりすましに限られない。なりすまし以外の他の不正は、後述の変形例で説明する。
- [0023] 例えば、第1サービスは、不正を検知するための不正検知エンジンが既に導入されているものとする。不正検知エンジンは、不正検知で利用されるプログラム又はシステムの総称である。第2サービスは、まだ不正検知エンジンが導入されておらず、ユーザからの通報を受けて人手で不正の判定が行われているものとする。また、第2サービスでは、なりすましが増えており、不正検知エンジンの導入が検討されているものとする。以降、第1サービスの不正検知エンジンを、第1不正検知エンジンと記載する。第2サービスの不正検知エンジンを、第2不正検知エンジンと記載する。
- [0024] 第1実施形態では、第2不正検知エンジンを1から作成すると非常に手間がかかるので、第1不正検知エンジンを流用して第2不正検知エンジンが作成される。第1不正検知エンジンをそのままコピーして第2不正検知エンジンとしてもよいが、第1サービス及び第2サービスは、不正検知に必要なデータ等が完全に一致するわけではないので、第2サービスに合うようにカスタマイズされる。
- [0025] 第2不正検知エンジンの流用元となる第1不正検知エンジン自体は、公知の種々のエンジンを利用可能である。第1実施形態では、学習モデル及びルール両方を含む第1不正検知エンジンを例に挙げる。以降、第1不正検知エンジンの学習モデル及びルールを、第1学習モデル及び第1ルールと記載する。第2不正検知エンジンの学習モデル及びルールを、第2学習モデル及び第2ルールと記載する。第1学習モデル及び第2学習モデルを区別しない時は、単に学習モデルと記載する。第1ルール及び第2ルールを区別しない時は、単にルールと記載する。

[0026] 学習モデルは、機械学習を利用したモデルである。学習モデルは、AI (Artificial Intelligence) と呼ばれることもある。機械学習自体は、公知の種々の方法を利用可能である。第1実施形態の機械学習は、深層学習及び強化学習を含む意味である。学習モデルは、教師有り機械学習、半教師有り機械学習、又は教師無し機械学習の何れであってもよい。例えば、学習モデルは、ニューラルネットワークであってもよい。ルール自体も、公知の種々のルールを利用可能である。例えば、ルールは、ユーザ特徴情報に基づいて判定可能な条件を含む。

[0027] ユーザ特徴情報は、ユーザの特徴に関する情報である。ユーザ特徴情報は、静的な情報であってもよいし、動的な情報であってもよい。静的な情報とは、予め登録された情報である。静的な情報は、ユーザが自分で変更しない限りは、原則として変わらない情報である。例えば、ユーザID、ユーザ名、性別、メールアドレス、年齢、生年月日、国籍、住所、又はこれらの組み合わせは、静的な情報に相当する。動的な情報とは、ユーザの行動に関する情報である。動的な情報は、ユーザがアクセスするたびに変わりうる情報である。例えば、ユーザ端末30の場所、不正検知が実行される時の時間、ユーザ端末30の識別情報、又はこれらの組み合わせは、動的な情報に相当する。

[0028] 図2には、あるユーザが第1サービスにログインする際に取得されるユーザ特徴情報の一例が示されている。図2のように、第1実施形態では、ドメイン固有言語の一種であるJSONで定義されたデータ形式のユーザ特徴情報を例に挙げる。データ形式自体は、公知の形式であってもよい。例えば、ユーザ特徴情報は、波括弧で囲われた部分に、具体的なデータの内容が記述される。ダブルクォーテーションで囲われた部分に、文字列が記述される。数値は、ダブルクォーテーションで囲われない。

[0029] 例えば、あるユーザがログイン画面G1から第1サービスへのログインを要求すると、第1サーバ10は、図2のようなユーザ特徴情報を取得する。図2の例では、ユーザ特徴情報は、「original」といった名前のデ

ータを含む。「original」は、更に、「userid」、「ipaddress」、及び「time」といった名前のデータを含む。これらの個々のデータは、ユーザの何らかの特徴を示すデータである。

[0030] 「userid」は、ユーザが入力したユーザIDを示す。「ipaddress」は、ユーザ端末30のIPアドレスを示す。「time」は、ログインが要求された日時を示す。「deviceid」は、ユーザのユーザ端末30を識別可能なデバイスIDを示す。IPアドレスは、場所に応じて変わることがあるが、デバイスIDは、場所に応じては変わらない。

[0031] 「name」、「posts」、「followers」、「following」、「gender」、「email」、及び「age」は、それぞれユーザIDに関連付けられたユーザ名、投稿数、フォロワー数、フォロワー数、性別、メールアドレス、及び年齢を示す。ユーザ特徴情報に含まれる情報は、任意の情報であってよく、図2の例に限られない。第1サービスでは、図2のようなユーザ特徴情報に基づいて、不正が検知される。

[0032] 図3は、第1サービスにおける不正検知の流れの一例を示す図である。S1～S5の処理は、第1不正検知エンジンの処理である。例えば、あるユーザが第1サービスにログインしようとする時（S1）、第1サーバ10は、ユーザ特徴情報を取得する（S2）。第1実施形態では、図2のようなユーザ特徴情報が取得された後に、不正検知で利用する情報を取得するための集計処理等が実行される。ユーザ特徴情報には、集計処理等の計算結果が次々と追加される。この計算自体は、公知の種々の方法を利用可能である。追加されるデータもJSONのデータ形式に沿ったものとなる。

[0033] 例えば、第1サーバ10は、「ipaddress」に基づいて、IPアドレスに対応する市町村を特定し、ユーザの普段の利用中心地からの距離を計算する。第1サーバ10は、ユーザ特徴情報に、この距離を「feature1」のデータ名で追加する。第1サーバ10は、「time」に基づいて、「userid」が示すユーザIDで現在の時間帯にログインされた回数を集計する。第1サーバ10は、ユーザ特徴情報に、この回数を「feat

ure 2」のデータ名で追加する。他にも例えば、第1サーバ10は、種々の計算を実行可能である。第1サーバは、ユーザ特徴情報に、計算結果を所定のデータ名で次々と格納する。これらの計算は、第1サーバ10以外の他のコンピュータで実行されてもよい。

[0034] 第1サーバ10は、第1学習モデルにユーザ特徴情報を入力し、第1学習モデルから出力されたスコアを取得する(S3)。第1サーバ10は、ユーザ特徴情報に基づいて、第1ルールに合致するか否かを判定する(S4)。図3のように、複数の第1ルールが存在してもよいし、単一の第1ルールのみが存在してもよい。第1サーバ10は、第1学習モデルのスコアと、第1ルールの判定結果と、を後述の履歴データベースDB2に格納する(S5)。第1サーバ10は、第1学習モデルのスコアと、第1ルールの判定結果と、が何れも不正を示すものでなければ、ログインを許可する。第1サーバ10は、第1学習モデルのスコアと、第1ルールの判定結果と、の少なくとも一方が不正を示すものであれば、ログインを許可しない。この場合、他のパスワードや生体認証等を利用した追加認証が要求されてもよい。

[0035] ユーザが第1サービスへのログインを要求してからある程度の期間(例えば、数週間~数ヶ月程度)が経過すると、不正であるか否かを確定するための確定タイミングが訪れる(S6)。この期間の長さは、全ユーザで共通であってもよいし、ユーザに応じた長さであってもよい。確定タイミングが訪れると、第1サービスの管理者は、ユーザ特徴情報やその他の情報を利用して、不正であるか否かを判定する。第1サーバ10は、不正であるか否かの確定結果を、履歴データベースに格納する(S7)。

[0036] 第1サービスでは、以上のような流れで不正が検知される。第1サービス及び第2サービスは、互いのサービス自体は異なるが、SNSという点では同種のサービスである。このため、第1サービスにおける不正ユーザの傾向と、第2サービスにおける不正ユーザの傾向と、が似ている可能性がある。例えば、第1サービスで不正をした不正ユーザが、同じユーザ端末30を利用して、同じ時間帯に同じIPアドレスで同じ場所から、第2サービスで不

正をしようとすることがある。更に、不正ユーザがなりすまそうとする正当ユーザの年齢やフォロワー数等が似ていることもある。

[0037] このように、第1サービスの不正の傾向と、第2サービスの不正の傾向と、が似ていれば、第1不正検知エンジンを第2不正検知エンジンに流用しても、十分に精度の高い不正検知が可能である。ただし、第1サービスの不正の傾向と、第2サービスの不正の傾向と、が異なる可能性もある。この場合、第1不正検知エンジンを流用して第2不正検知エンジンを作成しても、第2サービスの不正を検知できない可能性がある。即ち、第2サービスにおけるなりすましが可能になり、第2サービスのセキュリティを高めることができない可能性がある。

[0038] そこで、第1実施形態の不正検知システムSは、第2サービスで実際に発生した不正を、第1不正検知エンジンを流用して作成された第2不正検知エンジンで検知可能か否かを判定する。この第2不正検知エンジンは、第2サービスで実際に発生した不正を検知可能であると判定された場合に適用される。これにより、第2不正検知エンジンの作成を簡易化しつつ、第2サービスにおけるセキュリティを高めるようにしている。以降、第1実施形態の詳細を説明する。

[0039] [1-3. 第1実施形態で実現される機能]

図4は、第1実施形態で実現される機能の一例を示す機能ブロック図である。

[0040] [1-3-1. 第1サーバで実現される機能]

データ記憶部100は、記憶部12を主として実現される。ユーザID取得部101、ユーザ特徴情報取得部102、及び不正検知部103は、制御部11を主として実現される。

[0041] [データ記憶部]

データ記憶部100は、第1サービスにおける不正を検知するために必要なデータを記憶する。例えば、データ記憶部100は、第1サービスデータベースDB1と、履歴データベースDB2と、を記憶する。

[0042] 図5は、第1サービスデータベースDB1の一例を示す図である。第1サービスデータベースDB1は、第1サービスのユーザに関する情報が格納されたデータベースである。例えば、第1サービスデータベースDB1には、ユーザID、パスワード、ユーザ情報、利用状況情報、及び利用履歴情報が関連付けられて格納される。あるユーザが第1サービスの利用登録を完了すると、第1サービスデータベースDB1に新たなレコードが作成され、このユーザのユーザID等が格納される。

[0043] ユーザIDは、ユーザを識別可能なユーザ識別情報の一例である。このためユーザIDと記載した箇所は、ユーザ識別情報と読み替えることができる。ユーザ識別情報は、ユーザID以外の名前と呼ばれる情報であってもよい。例えば、ユーザ名、ユーザアカウント、又はログインIDと呼ばれる情報が、ユーザ識別情報に相当してもよい。メールアドレス又は電話番号等の情報がユーザ識別情報として利用されてもよい。パスワードは、ログインに必要な認証情報である。

[0044] ユーザ情報は、ユーザにより登録された情報である。例えば、ユーザ情報は、ユーザ名、性別、メールアドレス、及び年齢を含む。ユーザ情報は、国籍、電話番号、生年月日、郵便番号、住所、職業、年収、第1サービスと連携する他のサービスのユーザID、又は家族構成といった他の情報を含んでもよい。ユーザ情報は、ユーザを何らかの形で分類するための属性（例えば、年齢層や趣味など）を含んでもよい。ユーザ情報が示す個々の内容は、ユーザの特徴の1つである。

[0045] 利用状況情報は、第1サービスの利用状況に関する情報である。例えば、利用状況情報は、投稿数、フォロワー数、フォロー数、投稿内容、他のユーザからのコメント、ユーザ間のメッセージ、及び第1サービスの設定を含む。あるユーザが第1サービスを利用すると、このユーザの利用状況情報が更新される。あるユーザの利用状況情報が他のユーザの利用により更新されることもある。例えば、あるユーザの利用状況情報は、他のユーザによりフォローされたりコメントが入力されたりすると更新される。利用状況情報が示

す個々の内容は、ユーザの特徴の1つである。

[0046] 利用履歴情報は、第1サービスの利用履歴に関する情報である。利用履歴は、行動履歴ということもできる。利用状況情報は、現時点における利用状況に関する情報であるのに対し、利用履歴情報は、過去における利用状況に関する情報である。例えば、利用履歴情報は、過去のログインにおける時間、過去のログインで利用されたIPアドレス、過去のログインで利用されたユーザ端末30のデバイスID、及び過去のログインにおけるユーザの行動が格納される。利用履歴情報は、ログインが発生すると更新される。利用履歴情報が示す個々の内容は、ユーザの特徴の1つである。

[0047] 図6は、履歴データベースDB2の一例を示す図である。履歴データベースDB2は、第1サービスにおける不正検知の履歴が格納されたデータベースである。例えば、履歴データベースDB2には、ユーザ特徴情報、結果情報、及びステータス情報が格納される。第1サービスへのログインが発生すると、履歴データベースDB2に新たなレコードが作成され、このログイン時に実行された不正検知に関する情報が格納される。

[0048] ユーザ特徴情報の詳細は、先述した通りである。結果情報は、不正検知の結果に関する情報である。第1実施形態では、結果情報が不正であるか否か（不正の有無）を示す場合を説明するが、結果情報は、不正の疑いを示すスコアであってもよい。スコアが数値で表現される場合、スコアが高いほど不正の疑いが高い。逆に、スコアは、正当性を示すものであってもよい。この場合、スコアが低いほど不正の疑いが高い。スコアは、数値以外にもSランク、Aランク、Bランクといったような文字等で表現されてもよい。スコアは、不正の確率又は蓋然性ということもできる。

[0049] ステータス情報は、不正検知のステータスに関する情報である。第1実施形態では、確定タイミングが訪れた場合に、ログイン時に実行された不正検知の結果が確定するので、ステータス情報は、不正検知の結果が確定したか否かを示す。図6の「未確定」は、確定タイミングが訪れていないことを示す。図6の「確定」は、確定タイミングが訪れたことを示す。第1不正検知

エンジンが判定を誤ることもあるので、第1サービスの管理者により、未確定の結果情報が修正されることもある。

[0050] なお、データ記憶部100が記憶するデータは、上記の例に限られない。データ記憶部100は、任意のデータを記憶可能である。例えば、データ記憶部100は、第1不正検知エンジンを記憶する。第1実施形態では、第1不正検知エンジンは、第1学習モデル及び第1ルールを含むので、データ記憶部100は、学習済みの第1学習モデルと、第1ルールの内容を示すデータと、を記憶する。

[0051] 例えば、第1サーバ10は、実際に発生した不正におけるユーザ特徴情報に基づいて、第1学習モデルの訓練データを取得する。この訓練データは、第1サービスの管理者によって作成されてもよいし、公知の自動生成手法を利用して取得されてもよい。例えば、訓練データは、ユーザ特徴情報を入力部分とし、不正であるか否かを出力部分とするペアになる。第1サーバ10は、訓練データの入力部分が入力された場合に、訓練データの出力部分が出力されるように、第1学習モデルの学習を行う。第1学習モデルの学習方法自体は、公知の機械学習で利用されている方法を利用すればよい。第1サーバ10は、学習済みの第1学習モデルをデータ記憶部100に記録する。第1サーバ10は、教師無し学習又は半教師有り学習を利用して学習させた第1学習モデルをデータ記憶部100に記録してもよい。

[0052] 例えば、第1サーバ10は、第1サービスの管理者が作成した第1ルールを、データ記憶部100に記録する。第1ルールは、管理者が手動で作成しなくてもよく、公知のルール生成方法が利用されてもよい。例えば、決定木を作成する機械学習を利用して第1ルールが生成されてもよい。第1ルールは、不正を検知するためのものに限られず、正当性を検知するためのものであってもよい。例えば、ある条件に合致したユーザを正当と判定するための第1ルールであってもよい。

[0053] [ユーザID取得部]

ユーザID取得部101は、ユーザIDを取得する。このユーザIDは、

不正検知の対象となるユーザのユーザIDである。第1実施形態では、ログインが要求された場合に不正検知が実行されるので、ユーザID取得部101は、ログインが要求されたユーザIDを取得する。例えば、ユーザID取得部101は、ログイン画面G1の入力フォームF10に入力されたユーザIDを取得する。

[0054] [ユーザ特徴情報取得部]

ユーザ特徴情報取得部102は、第1サービスにおけるユーザの特徴に関するユーザ特徴情報を取得する。このユーザは、不正検知の対象となるユーザである。ユーザ特徴情報の詳細は、先述した通りである。第1実施形態では、JSONを一例とするドメイン固有言語に関するデータ形式のユーザ特徴情報が利用されるので、ユーザ特徴情報取得部102は、所定のドメイン固有言語に関するデータ形式のユーザ特徴情報を取得する。

[0055] 例えば、ユーザ特徴情報取得部102は、ユーザID取得部101により取得されたユーザIDに基づいて、ユーザ特徴情報を取得する。ユーザ特徴情報取得部102は、第1サービスデータベースDB1を参照し、このユーザIDに関連付けられたユーザ情報の全部又は一部に基づいて、ユーザ特徴情報を取得する。図2のデータ例であれば、ユーザ特徴情報取得部102は、このユーザIDに関連付けられたユーザ名、投稿数、フォロワー数、フォロー数、性別、メールアドレス、及び年齢を、「original」のユーザ特徴情報として取得する。

[0056] なお、ユーザ特徴情報に含まれる「name」等のデータ名と、第1サービスデータベースDB1のフィールド名と、は同じであってもよいし異なってもよい。これらの名前が異なる場合には、「name」等のデータ名と、第1サービスデータベースDB1のフィールド名と、の関係を示すデータがデータ記憶部100に記憶されているものとする。ユーザ特徴情報取得部102は、第1サービスデータベースDB1のうち、不正検知で利用するフィールドとして予め定められたフィールドの情報を取得し、当該フィールドのフィールド名に対応するデータ名を示す文字列の後に、当該取得した情

報が記述されるように、ユーザ特徴情報を取得する。

[0057] 例えば、ユーザ特徴情報取得部102は、ユーザ端末30から取得した情報に基づいて、ユーザ特徴情報を取得してもよい。第1実施形態では、ユーザ端末30は、第1サーバ10にログインを要求する場合に、自身のIPアドレスと、デバイスIDと、を送信する。ユーザ特徴情報取得部102は、「ipaddress」及び「deviceid」のデータ名を示す文字列の後に、ユーザ端末30から取得したIPアドレス及びデバイスIDが記述されるように、ユーザ特徴情報を取得する。

[0058] なお、デバイスIDは、ユーザ端末30の個体識別番号やSIMカードに格納されたIDであってもよいが、第1実施形態では、第1サーバ10により発行されたIDであってもよい。この場合、第1サーバ10は、あるユーザ端末30で第1サービスの利用登録が行われたり、何らかのユーザIDでログインが発生したりすると、このユーザ端末30を識別可能なデバイスIDを発行する。このユーザ端末30には、このデバイスIDが記録される。ユーザ端末30は、ログイン時に、第1サーバ10に当該デバイスIDを送信すればよい。

[0059] 例えば、ユーザ特徴情報取得部102は、ユーザ端末30からログインが要求された場合に、「time」のデータ名を示す文字列の後に、現在の日時が記述されるように、ユーザ特徴情報を取得する。現在の日時は、リアルタイムクロック又はGPS等を利用して取得されるようにすればよい。ユーザ特徴情報取得部102は、ユーザ端末30からログインが要求された場合に、「userid」のデータ名を示す文字列の後に、ログインが要求されたユーザIDが記述されるように、ユーザ特徴情報を取得する。

[0060] 第1実施形態では、ユーザ特徴情報取得部102は、集計処理等の計算によって、不正検知に利用する情報を取得する。例えば、ユーザ特徴情報取得部102は、ユーザID取得部101により取得されたユーザIDに関連付けられた利用履歴情報に基づいて、ユーザの利用中心地を計算する。利用中心地は、過去の全期間又は一部の期間における利用場所の位置の平均である

。ユーザ特徴情報取得部102は、ユーザ端末30のIPアドレスから推定した位置と、利用中心地と、の距離を計算する。ユーザ特徴情報取得部102は、「feature1」のデータ名を示す文字列の後に、この距離が記述されるように、ユーザ特徴情報を取得する。

[0061] 例えば、ユーザ特徴情報取得部102は、ユーザID取得部101により取得されたユーザIDに関連付けられた利用履歴情報に基づいて、現在の時間帯における過去の利用回数を計算する。更に、ユーザ特徴情報取得部102は、この利用履歴情報に基づいて、過去に利用されたデバイスID及びIPアドレスを特定する。ユーザ特徴情報取得部102は、「feature2」のデータ名を示す文字列の後に、当該計算された利用回数と、当該特定されたデバイスID及びIPアドレスと、が記述されるように、ユーザ特徴情報を取得する。

[0062] 以上のように、ユーザ特徴情報取得部102は、不正検知に利用する計算結果が次々と追加されるように、ユーザ特徴情報を取得する。ユーザ特徴情報の取得方法は、種々の方法を利用可能であり、第1実施形態の例に限られない。例えば、ユーザ特徴情報取得部102は、集計処理等の計算を実行することなく、ユーザ特徴情報を取得してもよい。他にも例えば、ユーザ特徴情報取得部102は、集計処理等の計算だけが含まれるように、ユーザ特徴情報を取得してもよい。

[0063] [不正検知部]

不正検知部103は、ユーザ特徴情報と、第1不正検知エンジンと、に基づいて、第1サービスにおける不正を検知する。例えば、不正検知部103は、ユーザ特徴情報に基づいて、第1不正検知エンジンの第1学習モデルからの出力を取得する。第1学習モデルは、ユーザ特徴情報が入力されると、ユーザ特徴情報に基づいて、不正の疑いを示すスコアを計算して出力する。第1学習モデルがニューラルネットワークである場合には、入力されたユーザ特徴情報が必要に応じて畳み込まれる。不正検知部103は、第1学習モデルから出力されたスコアを取得する。第1学習モデルは、スコアではなく

、不正であるか否かを示すラベルを出力してもよい。この場合、不正検知部103は、第1学習モデルから出力されたラベルを取得する。

[0064] 例えば、不正検知部103は、ユーザ特徴情報に基づいて、第1不正検知エンジンの第1ルールに含まれる条件が満たされるか否かを判定する。個々の条件には、この条件が満たされた場合の不正の有無の判定結果が関連付けられているものとする。即ち、個々の条件には、この条件が満たされることは、不正であることを意味するか、それとも正当であることを意味するか、が定義されている。

[0065] 図3の例であれば、不正検知部103は、ユーザ特徴情報に基づいて、第1ルール1-1が満たされるか否かを判定する。例えば、不正検知部103は、ユーザ特徴情報に含まれる利用中心地からの距離が50km以上であるか否かを判定する。不正検知部103は、この距離が50km以上であれば、不正であると判定する。不正検知部103は、この距離が50km未満であれば、不正ではないと判定する。

[0066] 不正検知部103は、ユーザ特徴情報に基づいて、第1ルール1-2が満たされるか否かを判定する。例えば、不正検知部103は、ユーザ特徴情報に含まれる時間帯ごとの利用回数が2回未満であるか否かを判定する。不正検知部103は、ユーザ特徴情報に含まれるデバイスIDが利用履歴情報に含まれているか否かを判定することによって、普段使用しないデバイスであるか否かを判定する。不正検知部103は、利用回数が2回未満の時間帯であり、かつ、普段使用しないデバイスであれば、不正であると判定する。不正検知部103は、利用回数が2回以上の時間帯である、又は、普段使用するデバイスであれば、正当であると判定する。

[0067] 不正検知部103は、ユーザ特徴情報に基づいて、第1ルール1-3が満たされるか否かを判定する。例えば、不正検知部103は、ユーザ特徴情報に含まれる投稿数が500以上であるか否かを判定する。不正検知部103は、ユーザ特徴情報に含まれるフォロワー数が1000人以上であるか否かを判定する。不正検知部103は、ユーザ特徴情報に含まれるIPアドレス

が利用履歴情報に含まれているか否かを判定することによって、初めてのIPアドレスであるか否かを判定する。不正検知部103は、投稿数が500以上であり、フォロワー数が1000人以上であり、かつ、初めてのIPアドレスであれば、不正であると判定する。不正検知部103は、投稿数が500未満である、又は、フォロワー数が1000人未満である、又は、使用したことのあるIPアドレスであれば、正当であると判定する。

[0068] 不正検知部103は、他の第1ルールについても同様にして、ユーザ特徴情報に基づいて、第1ルールが満たされるか否かを判定する。不正検知部103は、所定数以上の第1ルールに合致した場合に不正であると判定してもよい。不正検知部103は、第1学習モデルからのスコアが閾値以上である、又は、第1ルールにより不正との結果が得られた場合に、ログインしようとしているユーザが不正であると判定する。なお、不正検知部103は、第1学習モデルからのスコアが閾値以上であり、かつ、第1ルールにより不正との結果が得られた場合に、ログインしようとしているユーザが不正であると判定してもよい。更に、不正検知部103は、第1学習モデルからのスコアと、不正との結果が得られた第1ルールの数と、に基づいて、ログインしようとしているユーザが不正であるか否かを判定してもよい。

[0069] [1-3-2. 第2サーバで実現される機能]

データ記憶部200は、記憶部22を主として実現される。他の各機能は、制御部21を主として実現される。

[0070] [データ記憶部]

データ記憶部200は、第2サービスにおける不正を検知するために必要なデータを記憶する。例えば、データ記憶部200は、第2サービスデータベースDB3と、不正情報データベースDB4と、を記憶する。

[0071] 図7は、第2サービスデータベースDB3の一例を示す図である。第2サービスデータベースDB3は、第2サービスのユーザに関する情報が格納されたデータベースである。例えば、第2サービスデータベースDB3には、ユーザID、パスワード、ユーザ情報、利用状況情報、及び利用履歴情報が

関連付けられて格納される。これらの情報は、第2サービスに関するものである点で第1サービスデータベースDB1と異なるが、個々の情報の詳細は、第1サービスと同様であってよい。

[0072] 図8は、不正情報データベースDB4の一例を示す図である。不正情報データベースDB4は、第2サービスで実際に発生した不正に関する不正情報が格納されたデータベースである。不正情報は、不正の内容を示す情報である。不正情報は、第2サービスの不正ユーザのユーザ特徴情報を含む。ユーザ特徴情報の意味は、第1サービスと同様である。第1実施形態では、第2サービスのユーザ特徴情報も、JSON等のドメイン固有言語に関するデータ形式になる。ただし、第2サービスのユーザ特徴情報に含まれるデータ名と、第1サービスのユーザ特徴情報に含まれるデータ名と、は異なってもよい。これらのデータ名は異なるが、データが示すものが同じ又は類似していればよい。

[0073] 例えば、不正情報データベースDB4には、ユーザ特徴情報及び結果情報を含む不正情報が格納される。結果情報の意味は、第1サービスと同様である。第1実施形態では、不正情報は、第2サービスにおける不正ユーザのユーザ特徴情報を含むので、結果情報は、不正であることを示す。不正情報は、第2サービスにおける正当ユーザのユーザ特徴情報を含んでもよい。正当ユーザの特徴も不正検知で利用可能な情報であり、第2サービスで実際に発生した不正に関する情報の1つである。不正情報が第2サービスにおける正当ユーザのユーザ特徴情報を含む場合、結果情報は、正当であることを示す。

[0074] [エンジン取得部]

エンジン取得部201は、第1サービスで不正を検知するための第1不正検知エンジンに基づいて、第2サービスで不正を検知するための第2不正検知エンジンを取得する。第2不正検知エンジンは、第1不正検知エンジンの少なくとも一部に基づいて作成される。例えば、第2サービスの管理者が手動で第1不正検知エンジンをカスタマイズすることによって、第2不正検知

エンジンが作成されてもよい。この場合、エンジン取得部201は、第2サービスの管理者の端末から、第2不正検知エンジンを取得する。

[0075] 第1実施形態では、第2サービスの管理者が手動で第1不正検知エンジンをカスタマイズするのではなく、エンジン取得部201が、第1不正検知エンジンを自動的にカスタマイズすることによって、第2不正検知エンジンを取得する場合を説明する。エンジン取得部201は、予め定められた方法に基づいて、第1不正検知エンジンの全部又は一部の内容を変更することによって、第1不正検知エンジンをカスタマイズする。エンジン取得部201は、第1不正検知エンジンの一部の内容を削除することによって、第1不正検知エンジンをカスタマイズしてもよい。エンジン取得部201は、第1不正検知エンジンに機能を追加することによって、第1不正検知エンジンをカスタマイズしてもよい。エンジン取得部201は、これらの変更、削除、及び追加を組み合わせてもよい。

[0076] 例えば、第1不正検知エンジンでは、所定のドメイン固有言語に関するデータ形式のユーザ特徴情報が利用されるので、このデータ形式のデータが第2サービスに合うようにカスタマイズされる。エンジン取得部201は、第1不正検知エンジンで利用されるユーザ特徴情報に含まれるデータ名を、第2サービスで利用されるデータ名に変更する。エンジン取得部201は、当該変更されたデータ名を含むユーザ特徴情報であって、ドメイン固有言語に関するデータ形式のユーザ特徴情報が利用される第2不正検知エンジンを取得する。

[0077] 第1サービスにおけるデータ名と、第2サービスにおけるデータ名と、の関係を示すデータは、データ記憶部200に予め記憶されているものとする。これらのデータ名は、互いに意味するものが同じ又は類似するものとする。例えば、第1サービスで「userid」のデータ名で定義されるユーザIDが、第2サービスで「loginid」のデータ名で定義されているのであれば、第1サービスの「userid」のデータ名と、第2サービスの「loginid」のデータ名と、が関連付けられている。エンジン取得部

201は、第1不正検知エンジンに含まれる「userid」のデータ名の部分を、第2サービスの「loginid」のデータ名に変更することによって、第1不正検知エンジンをカスタマイズする。

[0078] 他にも例えば、第1サービスで「time」のデータ名で定義される時間が、第2サービスで「date/time」のデータ名で定義されているのであれば、第1サービスの「time」のデータ名と、第2サービスの「date/time」のデータ名と、が関連付けられている。エンジン取得部201は、第1不正検知エンジンに含まれる「time」のデータ名の部分を、第2サービスの「date/time」のデータ名に変更することによって、第1不正検知エンジンをカスタマイズする。その他にもデータ名の変更を必要とするデータがあれば、エンジン取得部201は、第2サービスに合うようなデータ名に変更することによって、第1不正検知エンジンをカスタマイズする。

[0079] 第1実施形態では、第2不正検知エンジンは、第2学習モデルを含まずに、第2ルールのみを含むものとする。エンジン取得部201は、第1不正検知エンジンに含まれる第1ルールに基づいて、第2不正検知エンジンに含まれる第2ルールを取得する。例えば、エンジン取得部201は、第1ルールに含まれるデータ名を、第2サービスのデータ名に変更することによって、第2ルールを作成する。

[0080] 図9は、第2不正検知エンジンの取得方法の一例を示す図である。例えば、図3で説明した「feature1」のデータ名の利用中心地からの距離が、第2サービスで「distance」のデータ名で定義されるのであれば、エンジン取得部201は、第1ルールの「feature1」のデータ名を「distance」に変更することによって、第2ルール2-1を作成する。第2ルール2-1のように、第1ルール1-1の閾値が変更されてもよい。この場合、閾値の変更方法（倍率や差分等）がデータ記憶部200に予め定義されているものとする。エンジン取得部201は、この定義に基づいて、第1ルール1-1の閾値を変更し、第2ルール2-1の閾値として

決定する。他の閾値についても、変更の必要があるものは同様にして変更される。

[0081] 例えば、図3で説明した「feature 2」のデータ名の利用回数及び普段使用するユーザ端末30のデバイスIDが、第2サービスで「feature A」及び「feature B」といった別々のデータ名で定義されるのであれば、エンジン取得部201は、第1ルールの「feature 2」のデータ名を、「feature A」及び「feature B」の2つのデータ名に変更することによって、第2ルールを作成する。その他にも、エンジン取得部201は、第2サービスに合うようなデータ名に変更することによって、第1ルールをカスタマイズして第2ルールを取得する。

[0082] [不正情報取得部]

不正情報取得部202は、第2サービスで実際に発生した不正に関する不正情報を取得する。第1実施形態では、不正情報取得部202は、第2サービスで実際に発生した複数の不正に対応する複数の不正情報を取得する。個々の不正情報は、不正情報データベースDB4に格納されているので、不正情報取得部202は、不正情報データベースDB4に格納された複数の不正情報を取得する。不正情報取得部202は、不正情報データベースDB4に格納された全ての不正情報を取得してもよいし、一部の不正情報のみを取得してもよい。

[0083] [判定部]

判定部203は、不正情報に基づいて、第2サービスにおける不正を第2不正検知エンジンで検知可能か否かを判定する。即ち、判定部203は、第2不正検知エンジンの精度を評価する。第2不正検知エンジンの精度が閾値以上であることは、第2サービスにおける不正を第2不正検知エンジンで検知可能であることを意味する。第2不正検知エンジンの精度が閾値未満であることは、第2サービスにおける不正を第2不正検知エンジンで検知可能ではないことを意味する。

[0084] 第1実施形態では、第2不正検知エンジンには第2ルールが含まれるので

、判定部203は、不正情報に基づいて、第2サービスにおける不正を第2ルールで検知可能か否かを判定する。例えば、判定部203は、複数の不正情報に基づいて、第2不正検知エンジンの正解率を計算する。判定部203は、m個の不正情報に基づいて、第2不正エンジンから出力されたm個の不正検知の結果を取得する。判定部203は、m個の結果のうち、不正との結果が得られた割合を、正解率として計算する。

[0085] 判定部203は、当該計算された正解率に基づいて、第2サービスにおける不正を第2不正検知エンジンで検知可能か否かを判定する。正解率が閾値（例えば、60%～90%程度）以上であることは、第2サービスにおける不正を第2不正検知エンジンで検知可能であることを意味する。正解率が閾値未満であることは、第2サービスにおける不正を第2不正検知エンジンで検知可能ではないことを意味する。なお、判定部203は、正解率以外の指標を利用して、第2不正検知エンジンで検知可能か否かを判定してもよい。例えば、適合率又は再現率が利用されてもよい。後述の変形例のように、学習モデルを含む第2不正検知エンジンであれば、Log Loss等が利用されてもよい。

[0086] [適用部]

適用部204は、第2サービスにおける不正を第2不正検知エンジンで検知可能であると判定された場合に、第2サービスに第2不正検知エンジンを適用する。第2サービスに第2不正検知エンジンを適用するとは、データ記憶部200に第2不正検知エンジンを記録することである。第2不正検知エンジンで不正検知を実行することは、第2サービスに第2不正検知エンジンを適用することに相当する。即ち、第2不正検知エンジンの実運用を開始することは、第2サービスに第2不正検知エンジンを適用することに相当する。第1実施形態では、第2不正検知エンジンには第2ルールが含まれるので、適用部204は、第2サービスにおける不正を第2ルールで検知可能であると判定された場合に、第2サービスに第2ルールを適用する。

[0087] [1-3-3. ユーザ端末で実現される機能]

データ記憶部300は、記憶部32を主として実現される。他の各機能は、制御部31を主として実現される。データ記憶部300は、第1サービス及び第2サービスの少なくとも一方を提供するために必要なデータを記憶する。表示制御部301は、種々の画面を表示部35に表示させる。受付部302は、操作部34から種々の操作を受け付ける。

[0088] [1-4. 第1実施形態で実行される処理]

図10は、第1実施形態で実行される処理の一例を示すフロー図である。この処理は、制御部11, 21, 31が記憶部12, 22, 32に記憶されたプログラムに従って動作することによって実行される。

[0089] ユーザ端末30は、ログイン画面G1の入力フォームF10, F11にユーザID及びパスワードが入力されてボタンB12が選択されると、第1サーバ10に、第1サービスへのログインを要求する(S100)。第1サーバ10は、ユーザ端末30からログインの要求を受信すると、第1サービスデータベースDB1に基づいて、ユーザ特徴情報を取得し(S101)、第1不正検知エンジンを利用して不正検知を実行する(S102)。なお、ユーザID及びパスワードの組み合わせが第1サービスデータベースDB1に存在しなければ、S101及びS102の処理は実行されず、第1サービスへのログインも実行されない。

[0090] S101では、第1サーバ10は、第1サービスデータベースDB1のうち、ログインが要求されたユーザIDが格納されたレコードを参照する。第1サーバ10は、このレコードのユーザ情報及び利用状況情報のうちの全部又は一部を、ユーザ特徴情報の静的な情報として取得する。第1サーバ10は、ユーザ端末30からのログインの要求に含まれるIPアドレス等の情報を、ユーザ特徴情報の動的な情報として取得する。第1サーバ10は、このレコードの利用履歴情報に基づいて、先述した集計処理等の計算を実行し、ユーザ特徴情報に次々と追加する。

[0091] S102では、第1サーバ10は、S101で取得したユーザ特徴情報に基づいて、第1学習モデルから出力されたスコアを取得する。第1サーバ1

0は、スコアが閾値以上の場合に、不正であると判定する。第1サーバ10は、S101で取得したユーザ特徴情報に基づいて、第1ルールに含まれる個々の条件を満たすか否かを判定する。第1サーバ10は、閾値以上の第1ルールが満たされた場合に、不正であると判定する。第1サーバ10は、第1学習モデル及び第1ルールの少なくとも一方の不正検知の結果が不正だった場合に、不正であると判定する。第1サーバ10は、第1学習モデル及び第1ルールの両方の不正検知の結果が不正ではなかった場合に、正当であると判定する。第1サーバ10は、これらの判定結果に基づいて、履歴データベースDB2を更新する。

[0092] S102において、不正が検知された場合（S102；不正）、第1サーバ10は、ユーザをログインさせず、ユーザ端末30に所定のエラーメッセージを表示させる（S103）。S102において、不正が検知されない場合（S102；正当）、第1サーバ10は、第1サービスにユーザをログインさせる（S104）。以降、ユーザは、第1サービスを利用する。ユーザが第1サービスを利用すると、第1サービスデータベースDB1に格納された利用状況情報及び利用履歴情報が更新される。第1サーバ10は、確定タイミングが訪れると、不正の有無を確定させる（S105）。S105では、第1サーバ10は、第1サービスの管理者の端末から、不正の有無の確定結果を取得し、履歴データベースDB2を更新する。

[0093] 第2サーバ20は、第1不正検知エンジンに基づいて、第2不正検知エンジンを取得する（S106）。第2不正検知エンジンの取得方法は、先述した通りである。第2サーバ20は、不正情報データベースDB4を参照して不正情報を取得する（S107）。第2サーバ20は、S107で取得した不正情報に基づいて、S106で取得した第2不正検知エンジンの正解率を取得する（S108）。

[0094] 第2サーバ20は、S108で取得した第2ルールの正解率が閾値以上であるか否かを判定する（S109）。正解率が閾値以上であると判定された場合（S109；Y）、第2サーバ20は、第2サービスに、第2不正検知

エンジンを適用する（S 1 1 0）。以降、第 1 サービスにおける S 1 0 0 ~ S 1 0 5 と同様の処理により、第 2 不正検知エンジンを利用して第 2 サービスの不正が検知される。正解率が閾値未満であると判定された場合（S 1 0 9 ; N）、第 2 サーバ 2 0 は、第 2 サービスに、第 2 不正検知エンジンを適用せずに、本処理は終了する。

[0095] 第 1 実施形態の不正検知システム S によれば、第 1 不正検知エンジンに基づいて第 2 不正検知エンジンを取得し、第 2 サービスにおける不正を第 2 不正検知エンジンで検知可能であると判定された場合に、第 2 サービスに第 2 不正検知エンジンを適用する。これにより、第 2 不正検知エンジンの作成を簡易化しつつ、第 2 サービスにおけるセキュリティが高まる。例えば、第 1 不正検知エンジンを流用して作成された第 2 不正検知エンジンの精度を検証したうえで、第 2 サービスに第 2 不正検知エンジンが適用されるので、第 2 サービスで実際に発生している不正を検知可能な精度の高い第 2 不正検知エンジンを第 2 サービスに適用できる。また、第 1 サービスに実際に適用されている第 1 不正検知エンジンを流用することによって、精度の高い第 2 不正検知エンジンとすることができる。

[0096] また、不正検知システム S は、不正情報に基づいて、第 2 サービスにおける不正を第 2 ルールで検知可能であると判定された場合に、第 2 サービスに第 2 ルールを適用する。これにより、第 2 ルールの作成を簡易化しつつ、第 2 サービスにおけるセキュリティが高まる。例えば、第 2 サービスの管理者が 1 から第 2 ルールを作成するといった手間が発生することを防止できる。

[0097] また、不正検知システム S は、第 2 不正検知エンジンの正解率に基づいて、第 2 サービスにおける不正を第 2 不正検知エンジンで検知可能か否かを判定する。これにより、第 2 サービスの不正検知の精度が高い第 2 不正検知エンジンを適用し、第 2 サービスにおけるセキュリティが高まる。

[0098] また、不正検知システム S は、所定のドメイン固有言語が利用される第 1 不正検知エンジンに基づいて、ドメイン固有言語が利用される第 2 不正検知エンジンを取得する。ドメイン固有言語を利用することによって、第 1 不正

検知エンジンを流用しやすくなるので、第2不正検知エンジンを作成しやすくなる。例えば、第1不正検知エンジンに入力されるユーザ特徴情報のデータ形式をある程度維持したまま第2不正検知エンジンを作成できるので、第2不正検知エンジンを作成する手間をより効果的に軽減できる。例えば、JSONのデータ形式のデータ名を変更したり、このデータ名のデータが示す数値の閾値を変更したりするだけで第2不正検知エンジンを作成できる。

[0099] [2. 第2実施形態]

次に、不正検知システムSの別実施形態である第2実施形態を説明する。第1実施形態では、第1サービスにおける第1不正検知エンジンを流用して第2不正検知エンジンが取得される場合を説明した。第2サービスに第2不正検知エンジンが適用されると、第2サービスにおける不正を検知できる。第2サービスの不正ユーザは、第1サービスでも不正をすることがある。第2サービスにおける不正検知の結果を第1サービスにフィードバックできれば、第1サービスにおける不正検知の精度が高まると考えられる。

[0100] このため、第2実施形態では、第2サービスにおける不正検知の結果を利用して、第1サービスにおける不正検知が行われる場合を説明する。なお、第1実施形態と同様の構成は、説明を省略する。また、第2実施形態における第2不正検知エンジンは、第1実施形態のように取得されなくてもよい。即ち、第2実施形態の不正検知システムSは、第1実施形態の不正検知システムSを前提としなくてもよい。第2実施形態の不正検知システムSは、第1実施形態で説明した機能を省略可能である。第2不正検知エンジンは、第1不正検知エンジンを流用せずに、第2サービスの管理者により作成されてもよい。

[0101] [2-1. 第2実施形態で実現される機能]

図11は、第2実施形態で実現される機能の一例を示す機能ブロック図である。

[0102] [2-1-1. 第1サーバで実現される機能]

データ記憶部100、ユーザID取得部101、及びユーザ特徴情報取得

部102は、第1実施形態と同様である。不正検知部103は、第1実施形態と共通の機能を有するが、一部の機能が異なる。結果情報取得部104は、制御部11を主として実現される。

[0103] [結果情報取得部]

結果情報取得部104は、ユーザIDに基づいて、不正を検知するための不正検知エンジンが第1サービスとは異なる第2サービスにおけるユーザの不正検知の結果に関する結果情報を取得する。このユーザIDは、第1サービスのユーザIDであってもよいし、第2サービスのユーザIDであってもよい。第1実施形態で説明したように、第2サービスの不正検知は、所定のドメイン固有言語のデータ形式のユーザ特徴情報に基づいて実行されるので、結果情報取得部104は、ドメイン固有言語のデータ形式のユーザ特徴情報を利用して実行された、第2サービスにおける不正検知の結果に関する結果情報を取得する。

[0104] 第2実施形態では、あるユーザの第1サービスのユーザIDと、このユーザの第2サービスのユーザIDと、の関係を示すデータがデータ記憶部100に記憶されているものとする。以降、第1サービスのユーザIDを第1ユーザIDと記載し、第2サービスのユーザIDを第2ユーザIDと記載する。これらを区別しない時は、単にユーザIDと記載する。

[0105] 結果情報取得部104は、ユーザID取得部101により取得された第1ユーザIDに関連付けられた第2ユーザIDに基づいて、結果情報を取得する。結果情報取得部104は、第2サーバ20に対し、第1ユーザIDに関連付けられた第2ユーザIDに関連付けられた結果情報を要求する。第2サーバ20は、この要求を受信すると、後述の履歴データベースDB5を参照し、この第2ユーザIDに関連付けられた結果情報を第1サーバ10に送信する。結果情報取得部104は、第2サーバ20により送信された結果情報を取得する。

[0106] なお、第2サーバ20のデータ記憶部200に、あるユーザの第1ユーザIDと、このユーザの第2ユーザIDと、が関連付けられていてもよい。こ

の場合、結果情報取得部104は、第2サーバ20に対し、ユーザID取得部101により取得された第1ユーザIDとともに、結果情報を要求する。第2サーバ20は、この要求を受信すると、この第1ユーザIDに関連付けられた第2ユーザIDを取得する。第2サーバ20は、後述の履歴データベースDB5を参照し、この第2ユーザIDに関連付けられた結果情報を、第1サーバ10に送信する。結果情報取得部104は、第2サーバ20により送信された結果情報を取得する。

[0107] また、あるユーザの第1ユーザIDと、このユーザの第2ユーザIDと、が同じであってもよい。即ち、第1サービスと第2サービスで共通のユーザIDが利用されてもよい。この場合、結果情報取得部104は、第2サーバ20に対し、ユーザID取得部101により取得されたユーザIDとともに、結果情報を要求すればよい。第2サーバ20は、後述の履歴データベースDB5を参照し、このユーザIDに関連付けられた結果情報を第1サーバ10に送信する。結果情報取得部104は、第2サーバ20により送信された結果情報を取得する。

[0108] [不正検知部]

図12は、第2実施形態における第1サービスの不正検知の一例を示す図である。図12のように、第2実施形態では、第1サービスにおけるユーザ特徴情報に、第2サービスにおける不正検知の結果を示すデータが追加される。図12の例では、「service2」のデータ名のデータは、第2サービスの結果情報である。第2実施形態では、ユーザ特徴情報の一部に第2サービスの結果情報が組み込まれる場合を説明するが、この結果情報は、ユーザ特徴情報に組み込まれなくてもよい。第1不正検知エンジンの第1学習モデルは、第2サービスにおける不正検知の結果を特徴量の1つとして不正検知を実行する。第1不正検知エンジンの第1ルールは、第2サービスにおける不正検知の結果が条件の1つになる。

[0109] 不正検知部103は、第1サービスにおけるユーザ特徴情報と、第2サービスにおける結果情報と、に基づいて、第1サービスにおける不正を検知す

る。例えば、不正検知部103は、第1サービスにおけるユーザ特徴情報と、第2サービスにおける第2ルールに基づく結果情報と、に基づいて、第1サービスにおける不正を検知する。不正検知部103は、第2サービスにおける結果情報を第1サービスにおけるユーザ特徴情報の一部に組み込み、当該結果情報を含むユーザ特徴情報に基づいて、第1サービスにおける不正を検知する。

[0110] 例えば、図12のように、第1ルールの条件の1つとして、第2サービスにおける結果情報が含まれる。例えば、不正検知部103は、第2サービスにおける結果情報が不正を示す場合、第1サービスでも不正と判定する。他にも例えば、結果情報がスコアを示す場合、不正検知部103は、第1不正検知エンジンの第1学習モデルが出力したスコアと、結果情報が示すスコアと、の合計値が閾値以上であるか否かを判定することによって、第1サービスにおける不正を検知してもよい。更に、第2サービスの結果情報を含む訓練データに基づいて第1学習モデルの学習を行い、不正検知部103は、結果情報を含むユーザ特徴情報を第1学習モデルに入力し、第1学習モデルからの出力を取得してもよい。不正検知部103が実行する個々の処理は、機械学習を利用した学習モデルによって、自動的に実行されてもよい。

[0111] [2-1-2. 第2サーバで実現される機能]

不正検知部205は、制御部21を主として実現される。不正検知部205は、第2不正検知エンジンに基づいて、第2サービスにおける不正を検知する。不正検知部205は、第2サービスにおける不正を検知するという点で第1サーバ10の不正検知部103と異なり、不正検知の方法自体は同様である。このため、不正検知部205の不正検知の処理は、不正検知部103の不正検知の処理の説明と同様である。

[0112] 第2サービスでは、第2サービスにおける不正に関する判定条件を含む第2ルールが利用される。不正検知部205の判定結果である結果情報は、第2ルールに基づいて判定された、第2サービスにおける不正検知の結果を示す。なお、エンジン取得部201、不正情報取得部202、判定部203、

及び適用部204は、第1実施形態と同様である。データ記憶部100は、第1実施形態と似ているが一部の機能が異なる。

[0113] 図13は、履歴データベースDB5の一例を示す図である。履歴データベースDB5は、第2サービスにおける不正検知の履歴が格納されたデータベースである。例えば、履歴データベースDB5には、ユーザ特徴情報、結果情報、及びステータス情報が格納される。第2サービスへのログインが発生すると、履歴データベースDB2に新たなレコードが作成され、このログイン時に実行された不正検知に関する情報が格納される。これらの情報が第2サービスに関するものである点で履歴データベースDB2と異なり、個々の情報の詳細は、第1サービスと同様である。ログインからある程度の時間が経過して確定タイミングが訪れると不正の有無が確定する点についても、第1サービスと同様であってよい。

[0114] [2-1-3. ユーザ端末で実現される機能]

ユーザ端末30の機能は、第1実施形態と同様である。

[0115] [2-2. 第2実施形態で実行される処理]

図14は、第1実施形態で実行される処理の一例を示すフロー図である。この処理は、制御部11, 21, 31が記憶部12, 22, 32に記憶されたプログラムに従って動作することによって実行される。

[0116] S200~S205の処理は、第1サービスにおける不正検知が第1サーバ10とユーザ端末30の間で実行される点と、第2サービスにおける不正検知が第2サーバ20とユーザ端末30の間で実行される点と、がS100~S105の処理と異なるが、他の点については、S100~S105と同様である。

[0117] S206及びS207の処理は、S100及びS101の処理と同様である。第1サーバ10は、ユーザ端末30からログインの要求を受信すると、第2サーバ20に対し、ログインが要求されたユーザの結果情報を要求する(S208)。第2サーバ20は、結果情報の要求を受信すると、履歴データベースDB5を参照して結果情報を取得し、第1サーバ10に対し、結果

情報を送信する（S209）。

[0118] 第1サーバ10は、結果情報を受信すると、S207で取得したユーザ特徴情報の一部として組み込み、第1不正検知エンジンを利用して不正検知を実行する（S210）。S210の処理は、第2サービスの結果情報が第1サービスの不正検知で考慮される点で第1実施形態とは異なるが、他の点については同様である。続くS211～S213の処理は、S103～S105の処理と同様である。なお、第1サービスの不正検知は実行されず、第1サービスへのログインも実行されない点は、第1実施形態と同様である。

[0119] 第2実施形態の不正検知システムSによれば、第1サービスにおけるユーザ特徴情報と、第2サービスにおける結果情報と、に基づいて、第1サービスにおける不正を検知する。これにより、第2サービスにおける不正検知の結果も総合的に考慮して第1サービスにおける不正検知が実行されるので、第1サービスにおけるセキュリティが高まる。例えば、第1サービスにおける不正の傾向と、第2サービスにおける不正の傾向と、が似ている場合には、第2サービスにおける結果情報は、第1サービスにおける不正検知でも参考になると考えられる。このため、第2サービスにおける結果情報を利用することによって、第1サービスの不正検知の精度が高まり、第1サービスにおけるセキュリティが高まる。

[0120] また、不正検知システムSは、第1サービスにおけるユーザ特徴情報と、第2サービスにおける第2ルールに基づく結果情報と、に基づいて、第1サービスにおける不正を検知する。これにより、第2サービスにおける第2ルールを利用した不正検知の結果も総合的に考慮して第1サービスにおける不正検知が実行されるので、第1サービスにおけるセキュリティが高まる。

[0121] また、不正検知システムSは、ドメイン固有言語のデータ形式のユーザ特徴情報を利用して実行された、第2サービスにおける不正検知の結果に関する結果情報を取得する。第1サービスと第2サービスで同じドメイン固有言語のデータ形式を共有することによって、第1サービスと第2サービスが連携しやすくなる。

[0122] [3. 変形例]

なお、本開示は、以上に説明した実施形態に限定されるものではない。本開示の趣旨を逸脱しない範囲で、適宜変更可能である。

[0123] [3-1. 第1実施形態に係る変形例]

まず、第1実施形態に係る変形例を説明する。図15は、第1実施形態に係る変形例の機能ブロック図である。変更部105は、制御部21を主として実現される。不正検知部205は、第2実施形態で説明した通りである。

[0124] [変形例1-1]

例えば、第1サービス及び第2サービスは、任意のサービスであってよく、第1実施形態及び第2実施形態の例に限られない。例えば、第1サービスは、第1電子決済サービスであり、第2サービスは、第1電子決済サービスとは異なる第2電子決済サービスであってもよい。以降の変形例では、これらの電子決済サービスにおける不正検知が実行される場合を説明する。なお、第1電子決済サービスを単に第1サービスと記載し、第2電子決済サービスを単に第2サービスと記載する。

[0125] 第1サービス及び第2サービスでは、任意の電子決済が可能であり、例えば、クレジットカード、デビットカード、電子マネー、ポイント、電子キャッシュ、銀行口座、仮想通貨、ウォレット、又はその他の電子バリューといった決済手段を利用可能である。バーコード又は二次元コードを利用した電子決済、近距離無線通信を利用した電子決済、又は生体認証を利用した電子決済も利用可能である。変形例1-1では、第1サービスでクレジットカードが利用され、第2サービスで電子マネーが利用される場合を例に挙げる。クレジットカードを利用した電子決済自体は、公知の種々の方法を利用可能である。電子マネーを利用した電子決済自体も、公知の種々の方法を利用可能である。

[0126] 変形例1-1の第1サービスデータベースDB1には、個々の決済に関する情報が格納される。第1サービスデータベースDB1は、決済データベースということもできる。例えば、第1サービスデータベースDB1には、カ

ード情報及び利用履歴情報が格納される。カード情報は、個々のクレジットカードに関する情報である。例えば、カード情報は、カード番号、有効期限、名義人情報、及び利用可能枠を含む。名義人情報は、名義人の氏名だけでなく、電話番号や住所等の情報が格納されてもよい。利用履歴情報は、クレジットカードの利用履歴に関する情報である。例えば、利用履歴情報は、利用日時、利用場所（利用店舗）、及び利用額を含む。カード番号により、個々のクレジットカードを識別可能なので、変形例 1-1 の第 1 サービスデータベース DB 1 には、ユーザ ID が含まれていなくてもよい。利用場所は、ユーザ端末 30 の GPS 情報を利用して取得されてもよいし、電子決済が実行された店舗の位置であってもよい。

[0127] 変形例 1-1 の履歴データベース DB 2 は、第 1 実施形態と同様であるが、ユーザ特徴情報は、SNS に関する情報ではなく、クレジットカードに関する情報を含む点で異なる。例えば、ユーザ特徴情報は、クレジットカードの利用場所、利用日時、利用額、利用中心地からの距離、普段利用する時間との違い、平均利用額との違い、キャッシングにおける借り入れ額、ユーザの年収といった情報が含まれる。結果情報及びステータス情報は、第 1 実施形態と同様である。

[0128] 変形例 1-1 の第 2 サービスデータベース DB 3 には、第 1 サービスデータベース DB 1 と同様、個々の決済に関する情報が格納される。例えば、第 1 サービスデータベース DB 1 には、電子マネー情報及び利用履歴情報が格納される。電子マネー情報は、個々の電子マネーに関する情報である。例えば、電子マネー情報は、電子マネー ID、残高、及び名義人情報を含む。利用履歴情報は、電子マネーの利用履歴に関する情報である。例えば、利用履歴情報は、利用日時、利用場所（利用店舗）、及び利用額を含む。電子マネー ID により、個々の電子マネーを識別可能なので、変形例 1-1 の第 2 サービスデータベース DB 3 には、ユーザ ID が含まれていなくてもよい。

[0129] 変形例 1-1 の不正情報データベース DB 4 は、第 1 実施形態と同様であるが、ユーザ特徴情報は、SNS に関する情報ではなく、電子マネーに関する

る情報を含む点で異なる。例えば、ユーザ特徴情報は、電子マネーの利用場所、利用日時、利用額、利用中心地からの距離、普段利用する時間との違い、平均利用額との違い、キャッシングにおける借り入れ額、ユーザの年収といった情報が含まれる。結果情報は、第1実施形態と同様である。履歴データベースDB5も、電子マネーに関する情報を含む点で、第2実施形態と異なるが、他の点は同様である。

[0130] 変形例1-1における不正検知の流れ自体は、図3を参照して説明した第1実施形態の流れと同様である。ただし、図3におけるS1のログインは、クレジットカードによる決済の要求になる。ユーザがクレジットカードを利用して決済を要求すると(S1)、第1サーバ10は、ユーザ特徴情報を取得する(S2)。例えば、第1サーバ10は、決済が要求されたクレジットカード番号、名義人情報、利用中心地からの距離、普段利用する時間との違い、平均利用額との違いといった種々の情報を取得し、JSON形式のユーザ特徴情報に次々と格納する。

[0131] 第1サーバ10は、ユーザ特徴情報に基づいて、第1学習モデルからの出力を取得する(S3)。第1学習モデルは、クレジットカードの利用に関するユーザ特徴情報と、クレジットカードを利用した不正の有無の情報と、を含む訓練データが学習されているものとする。第1学習モデルは、ユーザ特徴情報が入力されると、必要に応じて畳み込みを行い、不正の疑いを示すスコアを出力する。なお、クレジットカードの不正とは、不正ユーザによる他人のクレジットカードの利用である。ユーザ端末30に表示させたバーコード又は二次元コードを利用した電子決済であれば、他人のユーザID及びパスワードによるなりすましによってログインし、バーコード又は二次元コードを利用することが不正に相当する。

[0132] 第1サーバ10は、ユーザ特徴情報に基づいて、第1ルールの判定結果を取得する(S4)。第1ルールは、クレジットカードの利用に関するユーザ特徴情報の条件が含まれる。例えば、第1ルールは、第1実施形態と同様の利用中心地からの距離、普段利用する時間との違い、普段利用するデバイス

との違い以外にも、平均利用額との違い等のルールであってもよい。他にも例えば、第1ルールは、特定の名義人であれば正当と判定されるようなルールであってもよい。

[0133] 第1サーバ10は、第1学習モデルからの出力と、第1ルールの判定結果と、を履歴データベースDB2に格納する(S5)。第1サーバ10は、第1学習モデルから不正と判定された場合、又は、第1ルールで不正と判定された場合には、決済を許可しない。第1サーバ10は、第1学習モデルから不正と判定されず、かつ、第1ルールで不正と判定されない場合には、決済を実行する。決済から一定程度の期間が経過して確定タイミングが訪れると(S6)、第1サービスの管理者は、決済が不正であるか否かを確定し(S7)、履歴データベースDB2が更新される。

[0134] 変形例1-1でも、エンジン取得部201は、クレジットカードに関する不正を検知するための第1不正検知エンジンに基づいて、電子マネーに関する不正を検知するための第2不正検知エンジンを取得する。不正情報取得部202は、実際に電子マネーを利用して発生した不正に関する不正情報を取得する。判定部203は、電子マネーに関する不正を、第2不正検知エンジンで検知可能か否かを判定する。適用部204は、電子マネーに関する不正を第2不正検知エンジンで検知可能であると判定された場合に、第2サービスに第2不正検知エンジンを適用する。電子マネーの不正検知でも、クレジットカードの不正検知で利用される利用中心地からの距離、普段利用する時間との違い、普段利用するデバイスとの違いといったルール等を流用可能なことがあるので、その適正が検証される。

[0135] 変形例1-1によれば、第2不正検知エンジンの作成を簡易化しつつ、第2電子決済サービスにおけるセキュリティが高まる。

[0136] [変形例1-2]

例えば、第1実施形態で説明したように、第1不正検知エンジンでは、第1サービスにおける不正に関する第1スコアを出力する第1学習モデルが利用される。この場合、第1不正検知エンジンに含まれる第1ルールは、第1

スコアに関する条件を含んでもよい。例えば、不正と判定されることの条件として、第1スコアが閾値以上であることが設定されてもよい。逆に、正当と判定されることの条件として、第1スコアが閾値未満であることが設定されてもよい。

[0137] また、第2不正検知エンジンでは、第2サービスにおける不正に関する第2スコアを出力する第2学習モデルが利用される。この場合、第2不正検知エンジンに含まれる第2ルールは、第2スコアに関する条件を含んでもよい。例えば、不正と判定されることの条件として、第2スコアが閾値以上であることが設定されてもよい。逆に、正当と判定されることの条件として、第2スコアが閾値未満であることが設定されてもよい。第2ルールの閾値は、第1ルールの閾値と同じであってもよいし異なってもよい。第2スコアを示すデータ名や閾値は、第1実施形態と同様に、カスタマイズが行われてもよい。第2学習モデルは、後述の変形例1-3のように第1学習モデルを流用して作成されてもよいし、特に第1学習モデルを流用せずに作成されてもよい。

[0138] 変形例1-2によれば、第2不正検知エンジンに含まれる第2ルールには、第2サービスにおける不正に関する第2スコアを出力する第2学習モデルが利用される。これにより、第2学習モデルによる不正検知の結果も総合的に考慮して第2サービスの不正検知を実行し、第2サービスにおけるセキュリティが高まる。

[0139] [変形例1-3]

例えば、第2不正検知エンジンでは、第1学習モデルに基づく第2学習モデルが利用されてもよい。エンジン取得部201は、第1学習モデルのパラメータに基づくパラメータを含む第2学習モデルを取得する。第2学習モデルは、第1学習モデルと全く同じコピーの学習モデルであってもよいし、第1学習モデルの一部が第2サービスに合うように変更されてもよい。

[0140] エンジン取得部201は、第1学習モデルの一部を変更又は削除することによって第2学習モデルを取得してもよいし、第1学習モデルに機能を追加

することによって第2学習モデルを取得してもよい。例えば、第1学習モデルがニューラルネットワークである場合、エンジン取得部201は、第1学習モデルの入力層を、第2サービスのユーザ特徴情報に合うように変更した入力層に置き換えることによって、第2学習モデルを取得してもよい。

[0141] 例えば、第1サービスのユーザ特徴情報のうち、 n (n は自然数)個の項目が第1学習モデルに入力されるものとする。第2サービスのユーザ特徴情報のうち、 k (k は n 未満の自然数)個の項目が第2学習モデルに入力されるものとする。 k 個の項目が示す内容は、 n 個のうちの k 個の項目が示す内容と同じ又は似ているものとする。例えば、 k 個の項目として、第2サービスで電子マネーが利用された場合の利用中心地からの距離、利用額、利用時間といった内容が含まれる場合、 n 個の項目として、第1サービスでクレジットカードが利用された場合の利用中心地からの距離、利用額、利用時間といった内容が含まれるものとする。エンジン取得部201は、 n 個の項目が入力される第1学習モデルの入力層を、 k 個の項目のみになるように変更し、第2学習モデルを取得する。なお、 n 個の項目のうち足りない項目($n-k$ 個の項目)は、欠損値として扱われてもよい。

[0142] 例えば、エンジン取得部201は、第1学習モデルの出力層を、第2サービスの不正検知で得たい結果が得られるように変更した出力層に置き換えることによって、第2学習モデルを取得してもよい。例えば、第1学習モデルがスコアを出力するものとする。第2サービスでは、スコアではなく、不正であるか否かのラベルを得たいものとする。この場合、エンジン取得部201は、スコアを出力する第1学習モデルの出力層を、ラベルを出力する出力層に置き換えるように変更し、第2学習モデルを取得する。その他にも、エンジン取得部201は、第1学習モデルの中間層の一部を変更又は削除することによって第2学習モデルを取得してもよい。

[0143] 判定部203は、不正情報に基づいて、第2サービスにおける不正を第2学習モデルで検知可能か否かを判定する。この判定方法は、第1実施形態と同様に、正解率が利用されるようにすればよい。適用部204は、第2サー

ビスにおける不正を第2学習モデルで検知可能であると判定された場合に、第2サービスに第2学習モデルを適用する。第2サービスに適用されるのが第2ルールではなく第2学習モデルという点で第1実施形態と異なるだけであり、適用部204の処理自体は、第1実施形態で説明した通りである。

[0144] 変形例1-3によれば、判定部203は、不正情報に基づいて、第2サービスにおける不正を、第1学習モデルに基づく第2学習モデルで検知可能であると判定された場合に、第2サービスに第2学習モデルを適用する。これにより、第2学習モデルを作成する手間を省きつつ、セキュリティが高まる。

[0145] [変形例1-4]

例えば、不正検知システムSは、結果情報取得部104及び変更部105を含んでもよい。結果情報取得部104は、第2サービスに第2不正検知エンジンが適用された場合に、第2サービスにおける不正検知の結果に関する結果情報を取得する。結果情報取得部104は、第2実施形態で説明した通りである。

[0146] 変更部105は、結果情報に基づいて、第1不正検知エンジンを変更するための処理を実行する。この処理は、第2実施形態で説明したように、第2サービスの結果情報を、第1ルールの1つとして組み込む処理であってもよいし、他の処理であってもよい。他の処理としては、第1サービスの管理者に、第2サービスから取得された結果情報を通知し、第1不正検知エンジンの変更を促す処理であってもよい。他にも例えば、第2サービスの結果情報を、第1学習モデルに入力される特徴量の1つにする処理であってもよい。例えば、変更部105は、第1サービスの不正検知で結果情報を利用するように、第1不正検知エンジンを変更してもよい。

[0147] 変形例1-4によれば、第2サービスに第2不正検知エンジンが適用された場合に、第2サービスにおける不正検知の結果に関する結果情報に基づいて、第1不正検知エンジンを変更するための処理を実行する。これにより、第2サービスにおける不正検知の結果を第1サービスの不正検知にも利用し

、第1サービスにおけるセキュリティが高まる。

[0148] [変形例1-5]

例えば、第1実施形態で説明したように、第1サービスでは、第1サービスの第1ユーザが第1サービスを利用する場合に、第1不正検知エンジンに基づいて、当該第1ユーザの不正が検知される。第1サービスでは、第1ユーザが第1サービスを利用した後に、当該第1ユーザの不正の有無が確定する。

[0149] 第2サービスでは、第2サービスの第2ユーザが第2サービスを利用する場合に、第2不正検知エンジンに基づいて、当該第2ユーザの不正が検知されてもよい。第2サービスでは、第2ユーザが第2サービスを利用した後に、当該第2ユーザの不正の有無が確定してもよい。第2サービスは、第2サービスが利用されてから不正の有無が確定するまでの期間の長さが第1サービスよりも短いサービスであってもよい。即ち、第2サービスの利用から不正の確定タイミングまでの長さは、第1サービスの利用から不正の確定タイミングまでの長さよりも短い。

[0150] 変形例1-5によれば、第2サービスは、第2サービスが利用されてから不正の有無が確定するまでの期間の長さが第1サービスよりも短いサービスである。これにより、第2サービスにおける最新の不正の傾向を第1サービスにフィードバックできる。例えば、第1サービスで不正が確定するまでの期間を2月程度とし、第2サービスで不正が確定するまでの期間を2週間程度とする。この場合、不正の傾向が変化した時に、第1サービスでこの変化を把握するまでに2月程度かかることが考えられるが、より早く不正の傾向の変化を把握できる第2サービスの不正検知の結果を第1サービスにフィードバックすることにより、第1サービスにおける不正の変化に対応しやすくなる。

[0151] [変形例1-6]

例えば、第1不正検知エンジンでは、第3サービスにおける不正検知の結果が利用されてもよい。第3サービスは、第1サービス及び第2サービスと

は異なるサービスである。変形例 1-6 では、第 3 サービスが第 3 電子決済サービスである場合を説明するが、第 3 サービスは、他の任意のサービスであってよく、第 3 電子決済サービスに限られない。第 3 電子決済サービスで利用可能な決済手段は、任意の決済手段であってよく、変形例 1-6 では、ポイントである場合を説明する。ポイントを利用した電子決済自体は、公知の方法を利用可能である。

[0152] 例えば、第 1 不正検知エンジンの第 1 学習モデルは、第 3 サービスにおける不正検知の結果を、特徴量の 1 つとして利用する。第 1 不正検知エンジンの第 1 ルールは、第 3 サービスにおける不正検知の結果が条件の 1 つとして利用する。エンジン取得部 201 は、第 1 不正検知エンジンに基づいて、第 3 サービスにおける不正検知の結果が利用される第 2 不正検知エンジンを取得する。第 2 不正検知エンジンの第 2 学習モデルは、第 3 サービスにおける不正検知の結果を、特徴量の 1 つとして利用する。第 2 不正検知エンジンの第 2 ルールは、第 3 サービスにおける不正検知の結果が条件の 1 つとして利用する。変形例 1-6 でも、エンジン取得部 201 は、第 3 サービスにおける不正検知の結果を示すデータ名や閾値を、必要に応じて変更することによって、第 2 不正検知エンジンを取得してもよい。

[0153] 変形例 1-6 によれば、第 1 不正検知エンジンに基づいて、第 3 サービスにおける不正検知の結果が利用される第 2 不正検知エンジンを取得する。これにより、第 3 サービスにおける不正検知の結果を第 2 サービスに利用できるため、種々のサービスにおける不正検知の結果を総合的に考慮し、第 2 サービスのセキュリティが高まる。

[0154] [変形例 1-7]

例えば、複数の第 1 サービスが存在する場合に、エンジン取得部 201 は、複数の第 1 サービスに対応する複数の第 1 不正検知エンジンに基づいて、第 2 不正検知エンジンを取得してもよい。例えば、ある第 1 サービスは、あるカード会社が提供する電子決済サービスである。他の第 1 サービスは、他のカード会社が提供する電子決済サービスである。個々の第 1 不正検知エン

ジンの流用方法は、第1実施形態で説明した通りである。エンジン取得部201は、複数の第1不正検知エンジンのうちの全部を流用する必要はなく、その一部のみを流用してもよい。エンジン取得部201は、複数の不正検知エンジンの各々のデータ名や閾値を変更することによって、第2不正検知エンジンを取得する。

[0155] 変形例1-7によれば、複数の第1サービスに対応する複数の第1不正検知エンジンに基づいて、第2不正検知エンジンを取得する。これにより、複数の第1サービスを総合的に考慮して第2不正検知エンジンを取得し、第2サービスのセキュリティが高まる。

[0156] [変形例1-8]

エンジン取得部201は、複数の第1サービスのうち、第2サービスに関連付けられた第1サービスの第1不正検知エンジンに基づいて、第2不正検知エンジンを取得してもよい。第2サービスに関連付けられた第1サービスは、第2不正検知エンジンの流用元となる第1不正検知エンジンが適用されている第1サービスである。例えば、複数の第1サービスとして、クレジットカードの電子決済サービス、ポイントの電子決済サービス、電子決済サービス、旅行予約サービスが存在したとする。この中で、第2サービスである電子マネーの電子決済サービスの不正検知に有効な第1サービスが、クレジットカードの電子決済サービスと、ポイントの電子決済サービスと、の2つだったとすると、これら2つの第1サービスが第2サービスに関連付けられている。これらの関連付けは、データ記憶部200に予め記憶されているものとする。エンジン取得部201は、第2サービスに関連付けられていない第1サービスの第1不正検知エンジンは流用せずに、第2サービスに関連付けられた第1サービスの第1不正検知エンジンに基づいて、第2不正検知エンジンを取得する。

[0157] 変形例1-8によれば、複数の第1サービスのうち、第2サービスに関連付けられた第1サービスの第1不正検知エンジンに基づいて、第2不正検知エンジンを取得する。これにより、第2サービスと関連性の高い第1サービ

スの第1不正検知エンジンを流用できるので、第2サービスのセキュリティが高まる。

[0158] [変形例1-9]

例えば、第1実施形態と第2実施形態を組み合わせ、第1不正検知エンジンでは、結果情報が利用されてもよい。結果情報の利用方法自体は、第2実施形態で説明した通りである。変形例1-9では、第2実施形態で説明した各機能が実現される。

[0159] 変形例1-9によれば、第1不正検知エンジンで結果情報が利用される。これにより、第1サービスにおけるセキュリティが高まる。

[0160] [3-2. 第2実施形態に係る変形例]

次に、第2実施形態に係る変形例を説明する。図16は、第2実施形態に係る変形例の機能ブロック図である。確定情報取得部106及び変更判定部107は、制御部11を主として実現される。

[0161] [変形例2-1]

例えば、第2実施形態の不正検知システムSも、変形例1-1と同様、電子決済サービスに適用してもよい。第1サーバ10は、第2サービスにおける電子マネーの不正に関する結果情報を、第1サービスにおけるクレジットカードの不正検知に利用する。電子決済サービスに適用した場合の不正検知の方法は、変形例1-1で説明した通りである。以降の変形例も、電子決済サービスへの適用例を説明する。

[0162] 変形例2-1によれば、電子決済サービスにおけるセキュリティが高まる。

[0163] [変形例2-2]

例えば、第2実施形態で説明したように、第2サービスでは、複数の第2ルールが利用されてもよい。この場合、結果情報は、複数の第2ルールの各々に基づいて判定された、第2サービスにおける不正検知の結果を示す。即ち、結果情報は、複数の第2ルールが総合的に利用された不正検知の結果を示す。結果情報は、個々の第2ルールごとに不正の有無の判定結果が示され

ていてもよいし、不正と判定された第2ルールが1つでも存在するか否かが示されてもよい。

[0164] 不正検知部103は、第1サービスにおけるユーザ特徴情報と、第2サービスにおける複数の第2ルールに基づく結果情報と、に基づいて、第1サービスにおける不正を検知する。例えば、不正検知部103は、第2サービスにおける複数の結果情報をユーザ特徴情報に組み込み、第1サービスにおける不正を検知する。複数の結果情報が特徴量の1つとして考慮される点で第2実施形態とは異なるが、第1サービスにおける不正の検知自体は、第2実施形態と同様である。

[0165] 変形例2-2によれば、第1サービスにおけるユーザ特徴情報と、第2サービスにおける複数の第2ルールに基づく結果情報と、に基づいて、第1サービスにおける不正を検知する。複数の第2ルールが総合的に考慮されることによって、第1サービスにおけるセキュリティが高まる。

[0166] [変形例2-3]

例えば、第2サービスでは、第2サービスにおける不正に関する第2スコアを出力する第2学習モデルが利用されてもよい。第2学習モデルは、変形例1-3のようにして第1学習モデルが流用されることによって作成されてもよいし、特に第1学習モデルが流用されることなく作成されてもよい。この場合、結果情報は、第2学習モデルから出力された第2スコアに関する情報であってもよい。結果情報は、第2学習モデルが出力したスコアを示してもよいし、スコアが閾値以上であるか否かを示す情報（即ち、不正の有無を示す情報）であってもよい。

[0167] 不正検知部103は、第1サービスにおけるユーザ特徴情報と、第2サービスにおける第2スコアに基づく結果情報と、に基づいて、第1サービスにおける不正を検知する。第2スコアが特徴量の1つとして考慮される点で第2実施形態とは異なるが、第1サービスにおける不正の検知自体は、第2実施形態と同様である。第2実施形態で説明した第2ルールに基づく結果情報と同様に、第2スコアに基づく結果情報がユーザ特徴情報に組み込まれても

よいし、ユーザ特徴情報とは別に結果情報が利用されてもよい。例えば、第1ルールの一つとして、第2スコアが閾値以上であれば不正である、といったルールが存在してもよいし、第1学習モデルに入力される特徴量の一つとして利用されてもよい。

[0168] 変形例2-3によれば、第1サービスにおけるユーザ特徴情報と、第2サービスにおける第2スコアに基づく結果情報と、に基づいて、第1サービスにおける不正を検知する。これにより、第2サービスにおける不正検知の結果を第1サービスにも利用することによって、第1サービスにおけるセキュリティが高まる。

[0169] [変形例2-4]

例えば、第2サービスでは、複数の第2学習モデルが利用されてもよい。この場合、結果情報は、複数の第2学習モデルの各々から出力された第2スコアに関する情報である。即ち、結果情報は、複数の第2学習モデルが総合的に利用された不正検知の結果を示す。

結果情報は、個々の第2学習モデルごとの第2スコアが示されていてもよいし、第2スコアが閾値以上の第2学習モデルが1つでも存在するか否かが示されてもよい。

[0170] 不正検知部103は、第1サービスにおけるユーザ特徴情報と、第2サービスにおける複数の第2スコアに基づく結果情報と、に基づいて、第1サービスにおける不正を検知する。複数の第2スコアが考慮される点で第2実施形態とは異なるが、第1サービスにおける不正の検知自体は、第2実施形態と同様である。

[0171] 変形例2-4によれば、第1サービスにおけるユーザ特徴情報と、第2サービスにおける複数の第2スコアに基づく結果情報と、に基づいて、第1サービスにおける不正を検知する。これにより、複数の第2学習モデルが総合的に考慮されることによって、第1サービスにおけるセキュリティが高まる。

[0172] [変形例2-5]

例えば、複数の第2サービスが存在する場合、結果情報取得部104は、複数の第2サービスに対応する複数の結果情報を取得してもよい。例えば、ある第2サービスは、ある会社が提供する電子マネーの電子決済サービスである。他の第2サービスは、他の会社が提供する電子マネーの電子決済サービスである。複数の第2サービスが存在する点で第2実施形態とは異なるが、個々の結果情報自体は、第2実施形態と同様である。個々の第2サービスの第2サーバ20では、第1実施形態及び第2実施形態で説明した機能が実現される。

[0173] 不正検知部103は、第1サービスにおけるユーザ特徴情報と、複数の第2サービスに対応する複数の結果情報と、に基づいて、第1サービスにおける不正を検知する。複数の第2サービスの各々の結果情報が考慮される点で第2実施形態とは異なるが、第1サービスにおける不正の検知自体は、第2実施形態と同様である。不正検知部103は、複数の結果情報をユーザ特徴情報に組み込み、第1不正検知エンジンを利用して不正検知を実行する。例えば、不正検知部103は、結果情報が不正を示す第2サービスが1つでも存在すれば、第1サービスでも不正と判定してもよい。不正検知部103は、結果情報が不正を示す第2サービスの数が閾値以上であれば、第1サービスで不正と判定してもよい。更に、不正検知部103は、結果情報が不正を示す第2サービスの数を特徴量の1つとして、第1学習モデルに入力して第1学習モデルからの出力を取得してもよい。この場合、第2サービスの数と不正の有無との関係が第1学習モデルに学習済みであるものとする。

[0174] 変形例2-5によれば、第1サービスにおけるユーザ特徴情報と、複数の第2サービスに対応する複数の結果情報と、に基づいて、第1サービスにおける不正を検知する。これにより、複数の第2サービスにおける不正検知の結果を総合的に考慮して、第1サービスにおけるセキュリティが高まる。

[0175] [変形例2-6]

例えば、結果情報取得部104は、複数の第2サービスのうち、第1サービスに関連付けられた第2サービスに対応する結果情報を取得する。第1サ

ービスに関連付けられた第2サービスは、不正検知の結果を参考にする第2サービスである。例えば、複数の第2サービスとして、電子マネーの電子決済サービス、ポイントの電子決済サービス、電子決済サービス、旅行予約サービスが存在したとする。この中で、第1サービスであるクレジットカードの電子決済サービスの不正検知に有効な第2サービスが、電子マネーの電子決済サービスと、ポイントの電子決済サービスと、の2つだったとすると、これら2つの第2サービスが第1サービスに関連付けられている。これらの関連付けは、データ記憶部100に予め記憶されているものとする。結果情報取得部104は、第1サービスに関連付けられていない第2サービスに対応する結果情報は取得しない。第1サービスに関連付けられた第2サービスの結果情報が取得される点で第2実施形態とは異なるが、個々の結果情報自体は、第2実施形態と同様である。

[0176] 不正検知部103は、第1サービスにおけるユーザ特徴情報と、第1サービスに関連付けられた第2サービスに対応する結果情報と、に基づいて、第1サービスにおける不正を検知する。第1サービスに関連付けられた第2サービスの各々の結果情報が考慮される点で第2実施形態とは異なるが、第1サービスにおける不正の検知自体は、第2実施形態と同様である。不正検知部103は、第1サービスに関連付けられていない第2サービスに対応する結果情報は、第1サービスの不正検知で利用しない。

[0177] 変形例2-6によれば、第1サービスにおけるユーザ特徴情報と、第1サービスに関連付けられた第2サービスに対応する結果情報と、に基づいて、第1サービスにおける不正を検知する。これにより、第1サービスと関連性の高い第2サービスに対応する結果情報を利用して不正検知を実行できるので、第1サービスのセキュリティが高まる。

[0178] [変形例2-7]

例えば、変形例1-5と同様に、第2サービスは、第2サービスが利用されてから不正の有無が確定するまでの期間の長さが第1サービスよりも短いサービスであってもよい。

[0179] 変形例 2-7 によれば、第 2 サービスは、第 2 サービスが利用されてから不正の有無が確定するまでの期間の長さが第 1 サービスよりも短いサービスである。これにより、第 2 サービスにおける最新の不正の傾向を第 1 サービスにフィードバックできる。例えば、変形例 1-5 と同様の理由で、第 1 サービスにおける不正の変化に対応しやすくなる。

[0180] [変形例 2-8]

例えば、不正検知システム S は、変更部 105、確定情報取得部 106、及び変更判定部 107 を含んでもよい。確定情報取得部 106 は、第 1 サービスにおける不正の確定結果に関する確定情報を取得する。履歴データベース DB 2 に格納されたユーザ特徴情報、結果情報、及びステータス情報のデータセットは、確定情報の一例である。確定情報取得部 106 は、履歴データベース DB 2 を参照し、確定情報を取得する。確定情報取得部 106 は、履歴データベース DB 2 に格納された全部又は一部の確定情報を取得する。

[0181] 変更判定部 107 は、第 1 サービスにおける確定情報と、第 2 サービスにおける結果情報と、に基づいて、第 1 サービスで不正を検知するための第 1 不正検知エンジンを変更するか否かを判定する。例えば、変更判定部 107 は、第 1 サービスにおける確定情報で不正が確定したユーザが、第 2 サービスでも不正と判定されているか否かを判定する。変更判定部 107 は、第 1 サービスにおける確定情報で不正が確定したユーザが、第 2 サービスでも不正と判定されていると判定された場合に、第 1 不正検知エンジンを変更すると判定する。この場合、第 1 サービスの不正の傾向と、第 2 サービスの不正の傾向と、が似てきたので、第 2 サービスの結果情報を第 1 不正検知エンジンで利用すると判定される。

[0182] なお、変更判定部 107 は、所定数の確定情報のうち、第 2 サービスでも不正と判定された割合を計算してもよい。変更判定部 107 は、この割合が閾値以上である場合に、第 1 不正検知エンジンを変更すると判定してもよい。変更部 105 は、第 1 不正検知エンジンを変更すると判定された場合に、第 2 不正検知エンジンに基づいて、第 1 不正検知エンジンを変更するための

処理を実行する。この処理は、変形例 1-4 で説明した通りである。変更部 105 は、第 1 不正検知エンジンを変更すると判定されない場合には、この処理は実行しない。

[0183] 変形例 2-8 によれば、第 1 サービスにおける確定情報と、第 2 サービスにおける結果情報と、に基づいて、第 1 不正検知エンジンを変更すると判定された場合に、第 2 不正検知エンジンに基づいて、第 1 不正検知エンジンを変更するための処理を実行する。これにより、第 2 サービスにおける不正検知の結果を利用すると有効な場合に第 1 サービスの不正検知にも利用し、第 1 サービスにおけるセキュリティが高まる。

[0184] [変形例 2-9]

例えば、結果情報取得部 104 は、現時点から所定期間内における不正検知の結果に関する結果情報を取得してもよい。この期間は、任意の長さであってよく、例えば、数週間程度であってもよいし、数ヶ月程度であってもよい。結果情報取得部 104 は、所定期間よりも前に実行された不正検知の結果に関する結果情報は取得しない。履歴データベース DB5 には、第 2 サービスにおける不正検知が実行された日時が格納されているものとする。この日時が所定期間内の結果情報だけが取得される。

[0185] 不正検知部 103 は、第 1 サービスにおけるユーザ特徴情報と、所定期間内における第 2 サービスにおける結果情報と、に基づいて、ユーザの不正を検知する。所定期間外の第 2 サービスにおける結果情報が不正検知で利用されない点で第 2 実施形態と異なり、他の点については第 2 実施形態と同様である。

[0186] 変形例 2-9 によれば、不正検知部 103 は、第 1 サービスにおけるユーザ特徴情報と、所定期間内における第 2 サービスにおける結果情報と、に基づいて、ユーザの不正を検知する。これにより、比較的新しい不正検知の結果をフィードバックし、最新の不正の傾向に対応することができるので、第 1 サービスにおけるセキュリティが高まる。

[0187] [変形例 2-10]

例えば、第1実施形態及び第2実施形態を組み合わせ、結果情報取得部104は、適用部204により適用された第2不正検知エンジンに基づく不正検知の結果に関する結果情報を取得してもよい。

[0188] 変形例2-10によれば、第2不正検知エンジンに基づく不正検知の結果に関する結果情報を取得する。これにより、第2不正検知エンジンの作成を簡易化しつつ、第1サービスにおけるセキュリティが高まる。

[0189] [3-3. その他の変形例]

例えば、上記説明した変形例を組み合わせてもよい。

[0190] 例えば、ドメイン固有言語は、JSON以外の任意の言語を利用可能である。ユーザ特徴情報は、マークアップ言語を利用して取得されてもよい。ユーザ特徴情報のデータ形式は、他の種々の形式であってよい。例えば、不正検知システムSは、電子商取引サービス、電子チケットサービス、金融サービス、又は通信サービスといったサービスにおける不正検知にも適用可能である。例えば、第1実施形態のようにして、ある会社が提供する第1電子商取引サービスにおける第1不正検知エンジンを、他の会社が提供する第2電子商取引サービスにおける第2不正検知エンジンに流用してもよい。第2実施形態のようにして、他の会社が提供する第2電子商取引サービスにおける第2不正検知エンジンの不正検知の結果を、ある会社が提供する第1電子商取引サービスにおける第1不正検知エンジンの不正検知で利用してもよい。

[0191] また例えば、第1サーバ10で実現されるものとして説明した機能は、他のコンピュータで実現されてもよいし、複数のコンピュータで分担されてもよい。第2サーバ20で実現されるものとして説明した機能は、他のコンピュータで実現されてもよいし、複数のコンピュータで分担されてもよい。例えば、データ記憶部100、200に記憶されるものとしたデータは、データベースサーバに記憶されていてもよい。

請求の範囲

- [請求項1] 第1サービスにおけるユーザの特徴に関するユーザ特徴情報を取得するユーザ特徴情報取得手段と、
前記ユーザを識別可能なユーザ識別情報を取得するユーザ識別情報取得手段と、
前記ユーザ識別情報に基づいて、不正を検知するための不正検知エンジンが前記第1サービスとは異なる第2サービスにおける前記ユーザの不正検知の結果に関する結果情報を取得する結果情報取得手段と、
、
前記第1サービスにおける前記ユーザ特徴情報と、前記第2サービスにおける前記結果情報と、に基づいて、前記第1サービスにおける不正を検知する不正検知手段と、
を含む不正検知システム。
- [請求項2] 前記第2サービスでは、前記第2サービスにおける不正に関する判定条件を含む第2ルールが利用され、
前記結果情報は、前記第2ルールに基づいて判定された、前記第2サービスにおける不正検知の結果を示し、
前記不正検知手段は、前記第1サービスにおける前記ユーザ特徴情報と、前記第2サービスにおける前記第2ルールに基づく前記結果情報と、に基づいて、前記第1サービスにおける不正を検知する、
請求項1に記載の不正検知システム。
- [請求項3] 前記第2サービスでは、複数の前記第2ルールが利用され、
前記結果情報は、前記複数の第2ルールの各々に基づいて判定された、前記第2サービスにおける不正検知の結果を示し、
前記不正検知手段は、前記第1サービスにおける前記ユーザ特徴情報と、前記第2サービスにおける前記複数の第2ルールに基づく前記結果情報と、に基づいて、前記第1サービスにおける不正を検知する、
、

請求項 2 に記載の不正検知システム。

[請求項4]

前記第 2 サービスでは、前記第 2 サービスにおける不正に関する第 2 スコアを出力する第 2 学習モデルが利用され、

前記結果情報は、前記第 2 学習モデルから出力された前記第 2 スコアに関する情報であり、

前記不正検知手段は、前記第 1 サービスにおける前記ユーザ特徴情報と、前記第 2 サービスにおける前記第 2 スコアに基づく前記結果情報と、に基づいて、前記第 1 サービスにおける不正を検知する、

請求項 1 ～ 3 の何れかに記載の不正検知システム。

[請求項5]

前記第 2 サービスでは、複数の前記第 2 学習モデルが利用され、

前記結果情報は、前記複数の第 2 学習モデルの各々から出力された前記第 2 スコアに関する情報であり、

前記不正検知手段は、前記第 1 サービスにおける前記ユーザ特徴情報と、前記第 2 サービスにおける前記複数の第 2 スコアに基づく前記結果情報と、に基づいて、前記第 1 サービスにおける不正を検知する、

、

請求項 4 に記載の不正検知システム。

[請求項6]

前記結果情報取得手段は、複数の前記第 2 サービスに対応する複数の前記結果情報を取得し、

前記不正検知手段は、前記第 1 サービスにおける前記ユーザ特徴情報と、前記複数の第 2 サービスに対応する前記複数の結果情報と、に基づいて、前記第 1 サービスにおける不正を検知する、

請求項 1 ～ 5 の何れかに記載の不正検知システム。

[請求項7]

前記結果情報取得手段は、前記複数の第 2 サービスのうち、前記第 1 サービスに関連付けられた前記第 2 サービスに対応する前記結果情報を取得し、

前記不正検知手段は、前記第 1 サービスにおける前記ユーザ特徴情報と、前記第 1 サービスに関連付けられた前記第 2 サービスに対応す

る前記結果情報と、に基づいて、前記第1サービスにおける不正を検知する、

請求項1～6の何れかに記載の不正検知システム。

[請求項8]

前記第1サービスでは、前記第1サービスの第1ユーザが前記第1サービスを利用する場合に、第1不正検知エンジンに基づいて、当該第1ユーザの不正が検知され、

前記第1サービスでは、前記第1ユーザが前記第1サービスを利用した後に、当該第1ユーザの不正の有無が確定し、

前記第2サービスでは、前記第2サービスの第2ユーザが前記第2サービスを利用する場合に、第2不正検知エンジンに基づいて、当該第2ユーザの不正が検知され、

前記第2サービスでは、前記第2ユーザが前記第2サービスを利用した後に、当該第2ユーザの不正の有無が確定し、

前記第2サービスは、前記第2サービスが利用されてから不正の有無が確定するまでの期間の長さが前記第1サービスよりも短いサービスである、

請求項1～7の何れかに記載の不正検知システム。

[請求項9]

前記不正検知システムは、

前記第1サービスにおける不正の確定結果に関する確定情報を取得する確定情報取得手段と、

前記第1サービスにおける前記確定情報と、前記第2サービスにおける前記結果情報と、に基づいて、前記第1サービスで不正を検知するための第1不正検知エンジンを変更するか否かを判定する変更判定手段と、

前記第1不正検知エンジンを変更すると判定された場合に、前記第2不正検知エンジンに基づいて、前記第1不正検知エンジンを変更するための処理を実行する変更手段と、

を更に含む請求項1～8の何れかに記載の不正検知システム。

[請求項10] 前記結果情報取得手段は、現時点から所定期間内における不正検知の結果に関する前記結果情報を取得し、

前記不正検知手段は、前記第1サービスにおける前記ユーザ特徴情報と、前記所定期間内における前記第2サービスにおける前記結果情報と、に基づいて、前記ユーザの不正を検知する、

請求項1～9の何れかに記載の不正検知システム。

[請求項11] 前記不正検知システムは、

前記第1サービスで不正を検知するための第1不正検知エンジンに基づいて、前記第2サービスで不正を検知するための第2不正検知エンジンを取得する不正検知エンジン取得手段と、

前記第2サービスで実際に発生した不正に関する不正情報を取得する不正情報取得手段と、

前記不正情報に基づいて、前記第2サービスにおける不正を前記第2不正検知エンジンで検知可能か否かを判定する判定手段と、

前記第2サービスにおける不正を前記第2不正検知エンジンで検知可能であると判定された場合に、前記第2サービスに前記第2不正検知エンジンを適用する適用手段と、

を含み、

前記結果情報取得手段は、前記第2不正検知エンジンに基づく不正検知の結果に関する前記結果情報を取得する、

請求項1～10の何れかに記載の不正検知システム。

[請求項12] 前記ユーザ特徴情報取得手段は、所定のドメイン固有言語に関するデータ形式の前記ユーザ特徴情報を取得し、

前記結果情報取得手段は、前記ドメイン固有言語のデータ形式のユーザ特徴情報を利用して実行された、前記第2サービスにおける不正検知の結果に関する前記結果情報を取得する、

請求項1～11の何れかに記載の不正検知システム。

[請求項13] 前記第1サービスは、第1電子決済サービスであり、

前記第2サービスは、前記第1電子決済サービスとは異なる第2電子決済サービスである、

請求項1～12の何れかに記載の不正検知システム。

[請求項14]

第1サービスにおけるユーザの特徴に関するユーザ特徴情報を取得するユーザ特徴情報取得ステップと、

前記ユーザを識別可能なユーザ識別情報を取得するユーザ識別情報取得ステップと、

前記ユーザ識別情報に基づいて、不正を検知するための不正検知エンジンが前記第1サービスとは異なる第2サービスにおける前記ユーザの不正検知の結果に関する結果情報を取得する結果情報取得ステップと、

前記第1サービスにおける前記ユーザ特徴情報と、前記第2サービスにおける前記結果情報と、に基づいて、前記第1サービスにおける不正を検知する不正検知ステップと、

を含む不正検知方法。

[請求項15]

第1サービスにおけるユーザの特徴に関するユーザ特徴情報を取得するユーザ特徴情報取得手段、

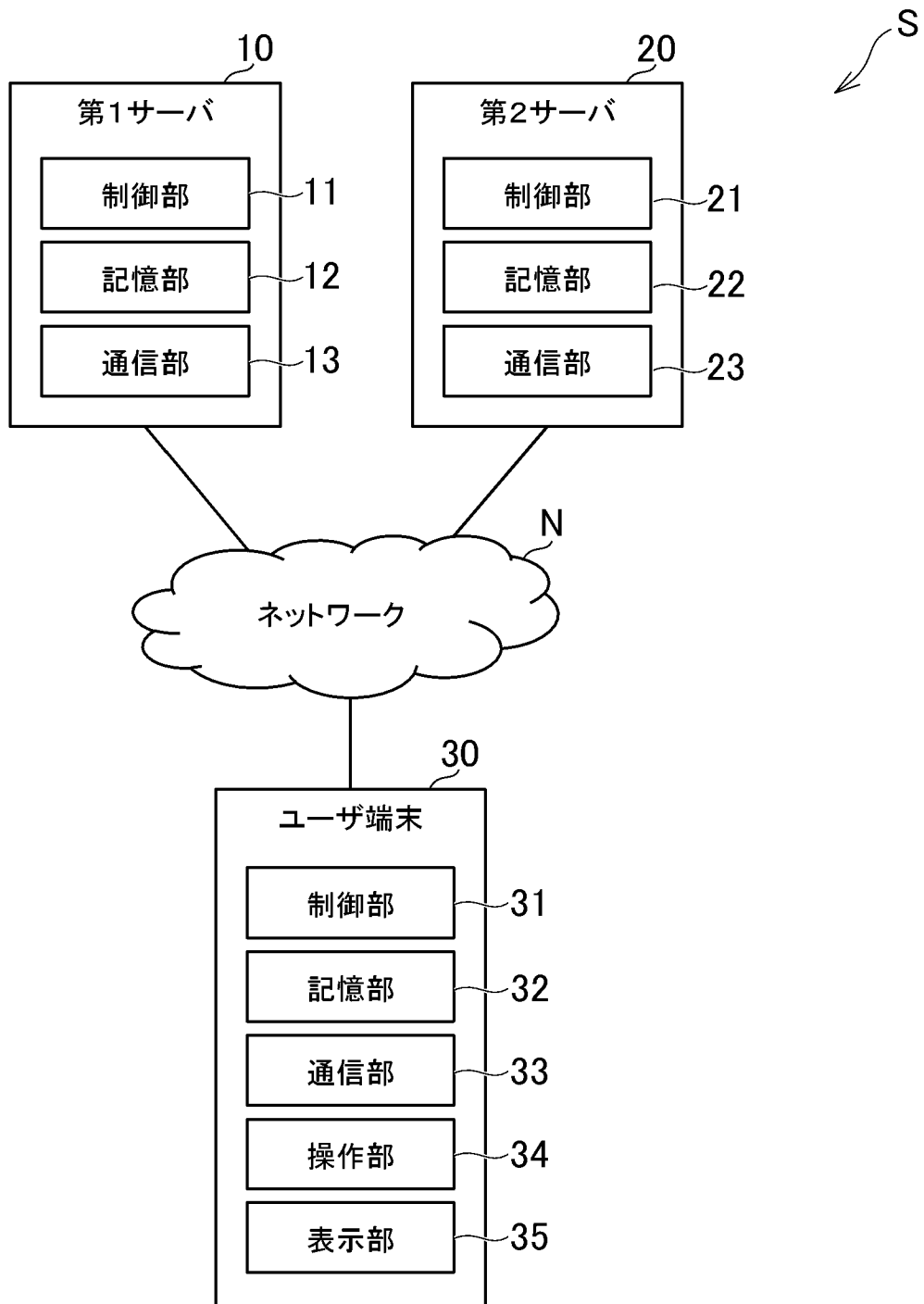
前記ユーザを識別可能なユーザ識別情報を取得するユーザ識別情報取得手段、

前記ユーザ識別情報に基づいて、不正を検知するための不正検知エンジンが前記第1サービスとは異なる第2サービスにおける前記ユーザの不正検知の結果に関する結果情報を取得する結果情報取得手段、

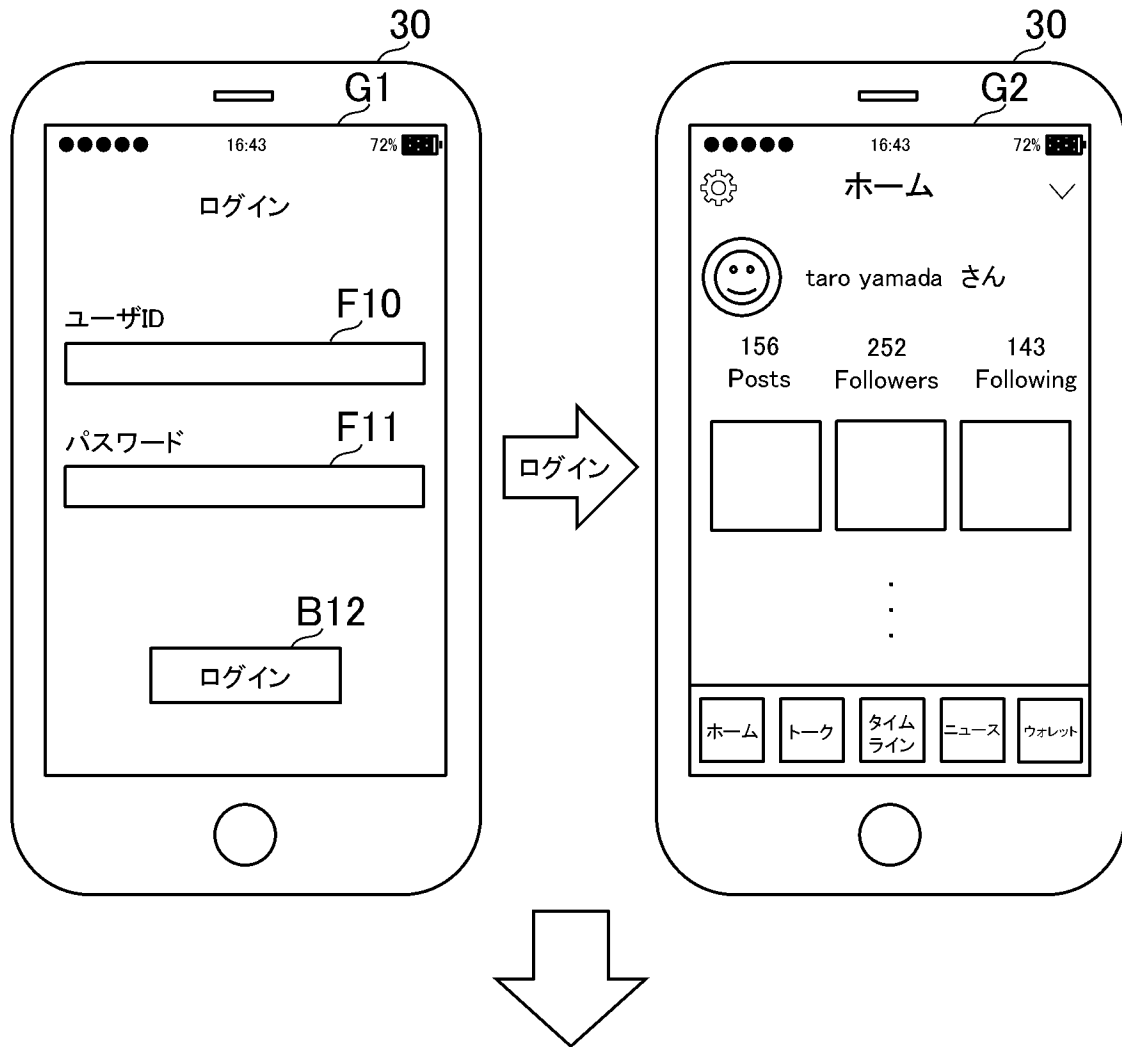
前記第1サービスにおける前記ユーザ特徴情報と、前記第2サービスにおける前記結果情報と、に基づいて、前記第1サービスにおける不正を検知する不正検知手段、

としてコンピュータを機能させるためのプログラム。

[図1]



[図2]



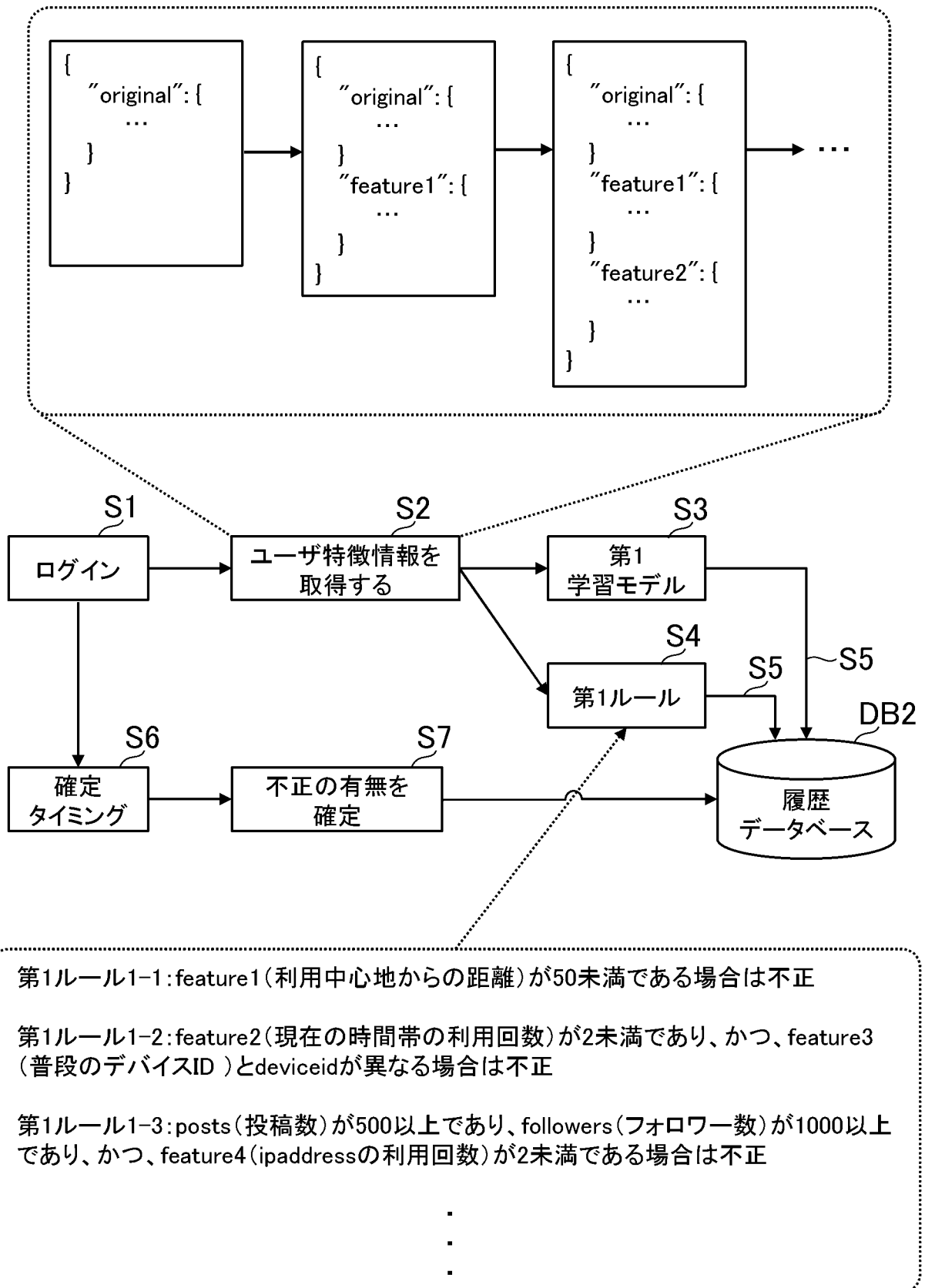
ユーザ特徴情報

```

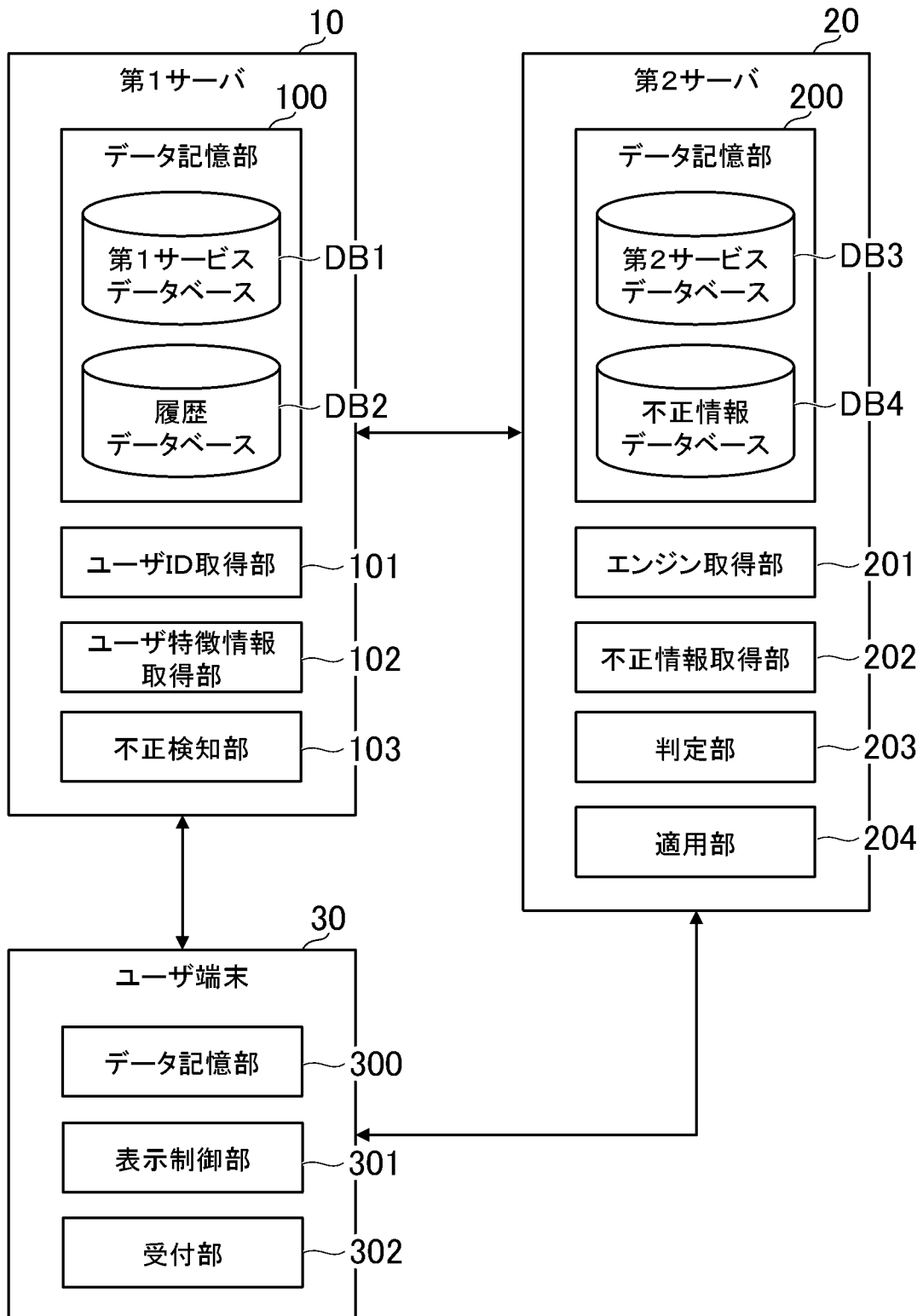
{
  "original": {
    "userid": "taro.yamada123",
    "ipaddress": "123.456.789.012",
    "time": "2021/8/20 16:43:14",
    "deviceid": "d00001",
    "name": "taro.yamada",
    "posts": 156,
    "followers": 252,
    "following": 143,
    "gender": "male",
    "email": "taro-yamada@abc.com",
    "age": 20,
    .
    .
    .
  }
}

```

[図3]



[図4]



[図5]

DB1

ユーザID	パスワード	ユーザ情報	利用状況情報	利用履歴情報
taro.yamada123	*****	ユーザ情報1-1	利用状況情報1-1	利用履歴情報1-1
hanako.kimura999	*****	ユーザ情報1-2	利用状況情報1-2	利用履歴情報1-2
hideo.tanaka001	*****	ユーザ情報1-3	利用状況情報1-3	利用履歴情報1-3
.
.
.

[図6]

DB2

ユーザ特徴情報	結果情報	ステータス情報
ユーザ特徴情報1-1	正当	確定
ユーザ特徴情報1-2	不正	確定
ユーザ特徴情報1-3	正当	未確定
.	.	.
.	.	.
.	.	.

[図7]

DB3

ユーザID	パスワード	ユーザ情報	利用状況情報	利用履歴情報
jiro.hakamada01	*****	ユーザ情報2-1	利用状況情報2-1	利用履歴情報2-1
humiko.ichikawa	*****	ユーザ情報2-2	利用状況情報2-2	利用履歴情報2-2
aki.yamaoka	*****	ユーザ情報2-3	利用状況情報2-3	利用履歴情報2-3
.
.
.

[図8]

DB4

不正情報	
ユーザ特徴情報	結果情報
ユーザ特徴情報2-1	不正
ユーザ特徴情報2-2	不正
ユーザ特徴情報2-3	不正
.	.
.	.
.	.

[図9]

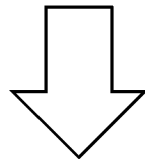
第1ルール

第1ルール1-1: feature1(利用中心地からの距離)が50未満である場合は不正

第1ルール1-2: feature2(現在の時間帯の利用回数)が2未満であり、かつ、feature3(普段のデバイスID)とdeviceidが異なる場合は不正

第1ルール1-3: posts(投稿数)が500以上であり、followers(フォロワー数)が1000以上であり、かつ、feature4(ipaddressの利用回数)が2未満である場合は不正

・
・
・



第1不正検知エンジンに含まれる第1ルールを
第2サービスに合うようにカスタマイズして
第2不正検知エンジンの第2ルールにする

第2ルール

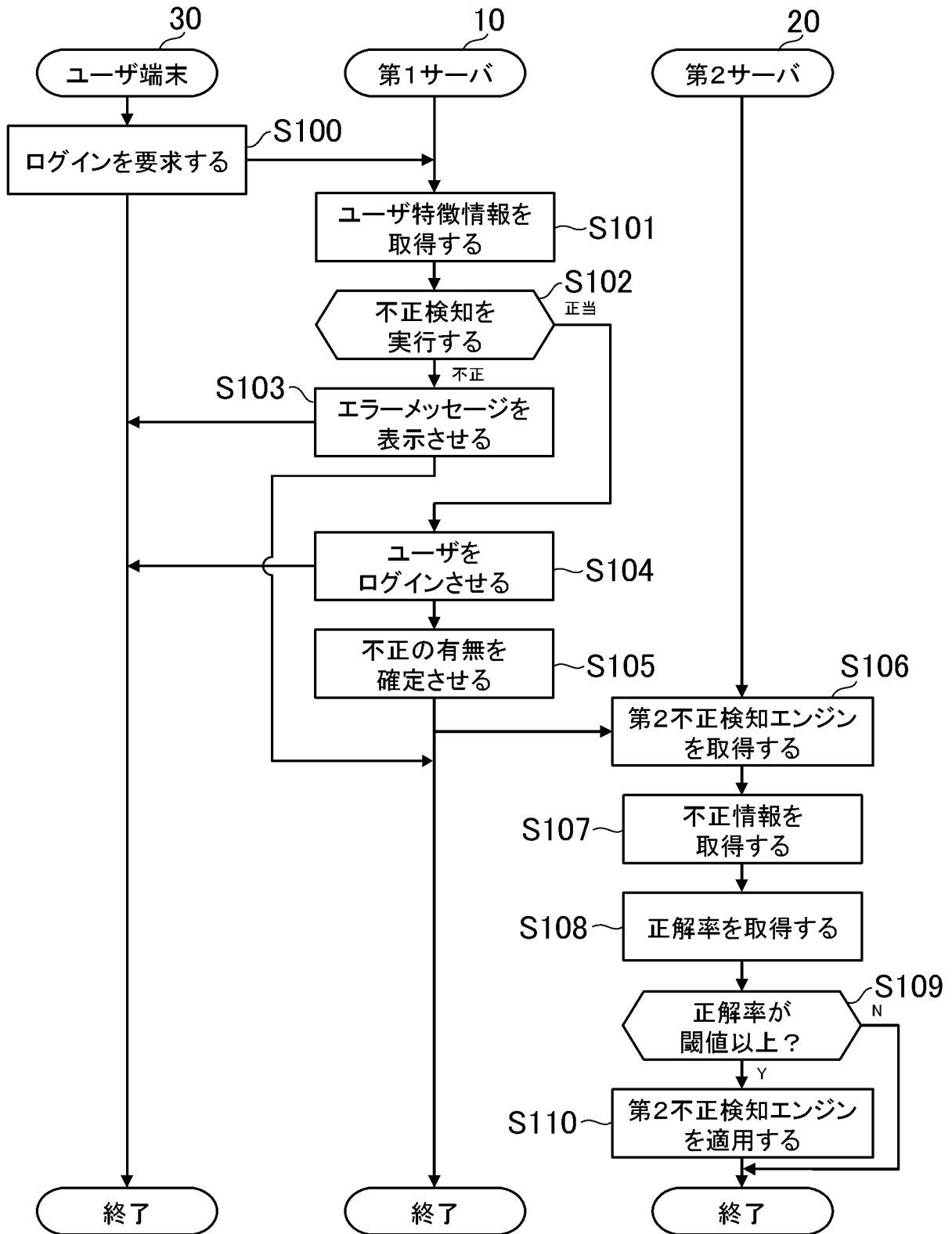
第2ルール2-1: distance(利用中心地からの距離)が100未満である場合は不正

第2ルール2-2: featureA(現在の時間帯の利用回数)が2未満であり、かつ、feature(普段のデバイスID)とterminalidが異なる場合は不正

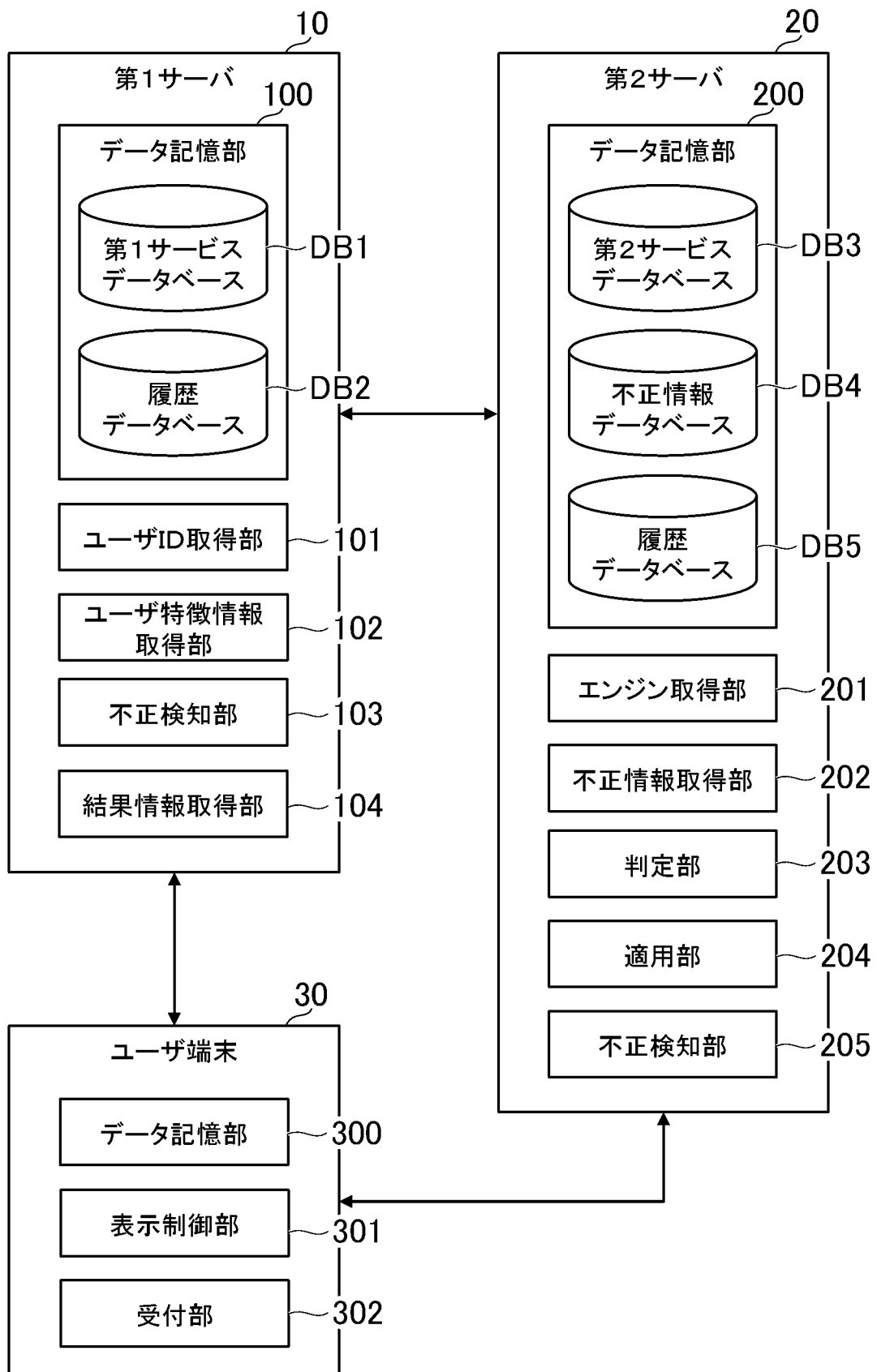
第2ルール2-3: messages(投稿数)が300以上であり、followers(フォロワー数)が500以上であり、かつ、feature4(ipaddressの利用回数)が2未満である場合は不正

・
・
・

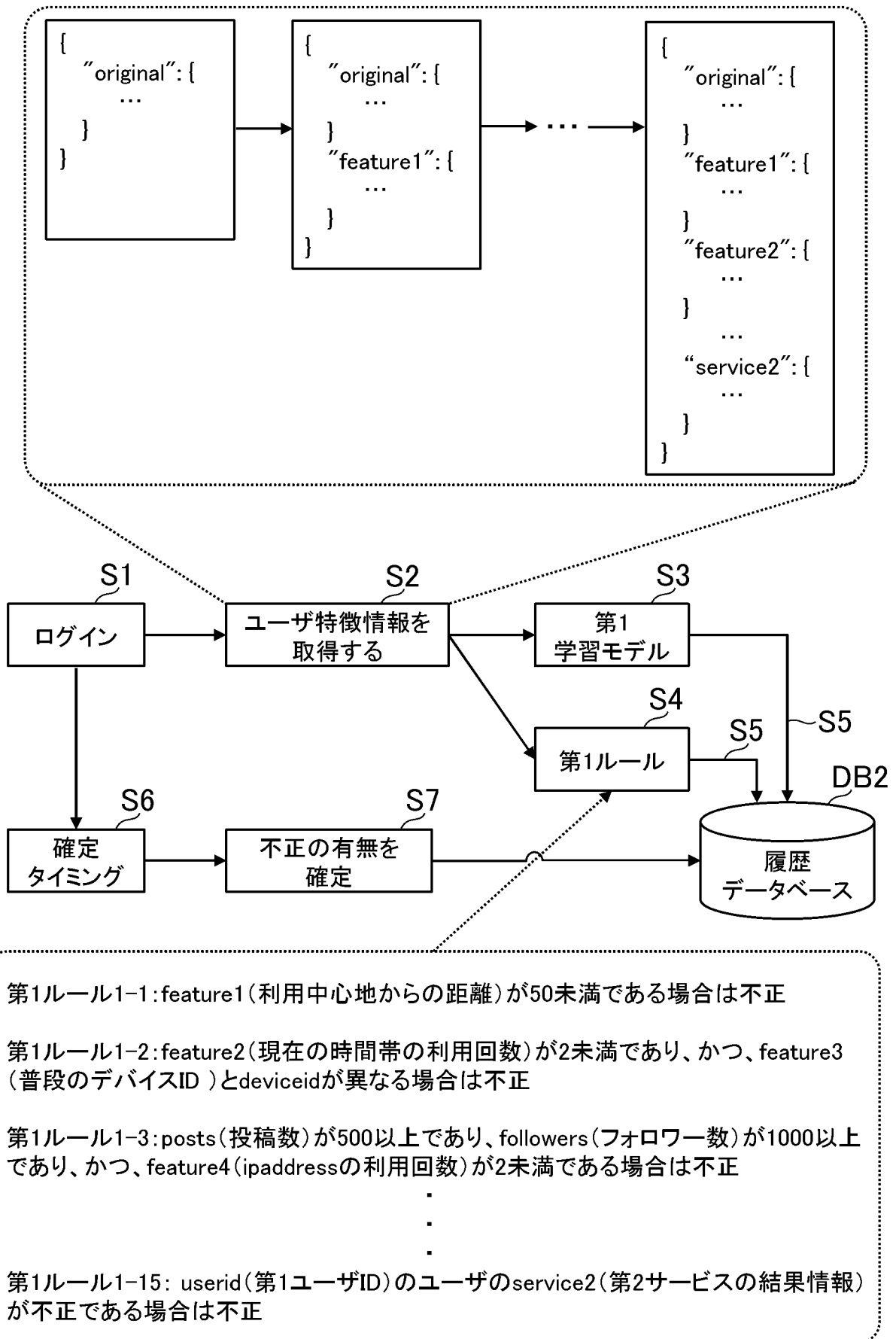
[図10]



[図11]



[図12]

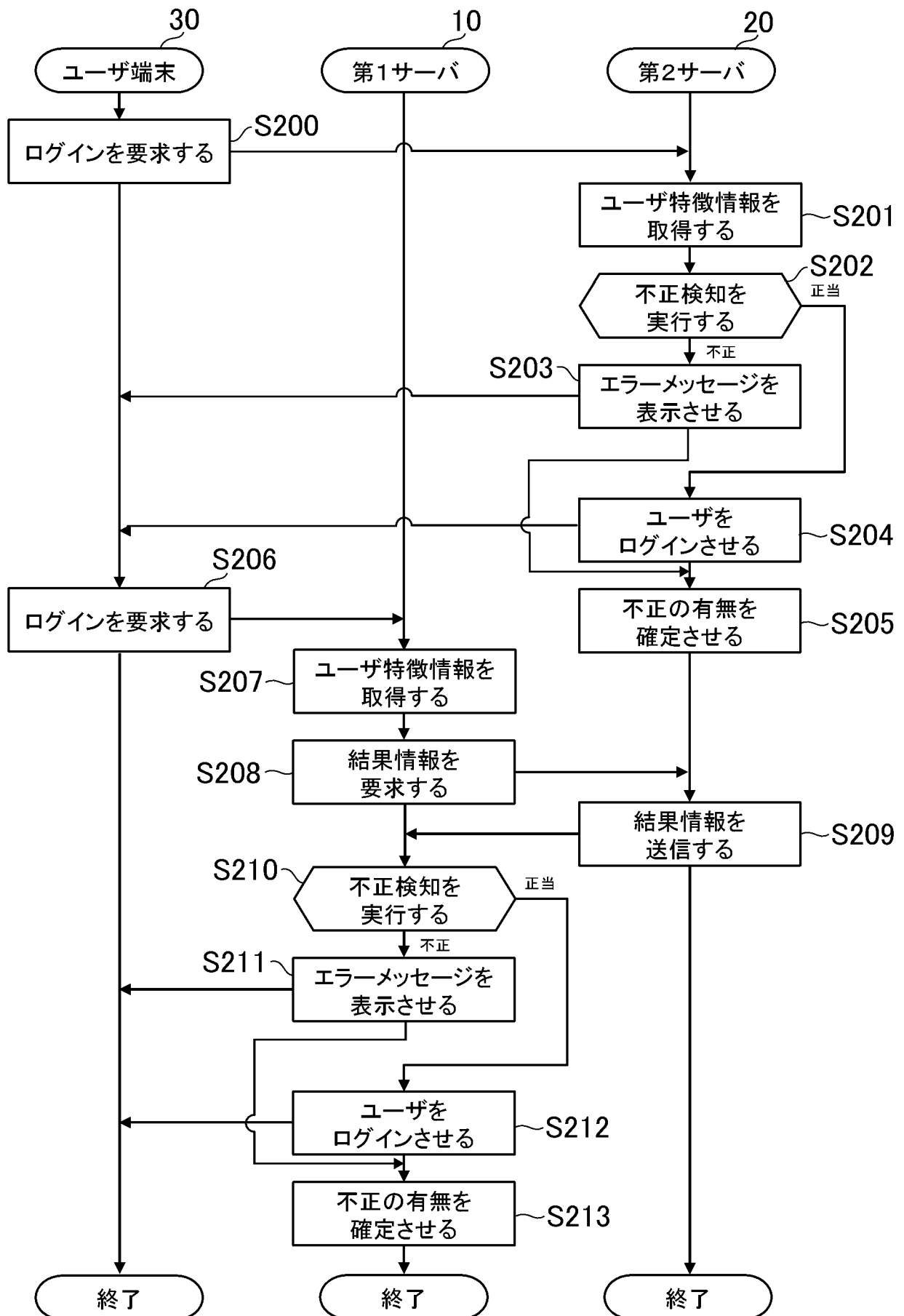


[図13]

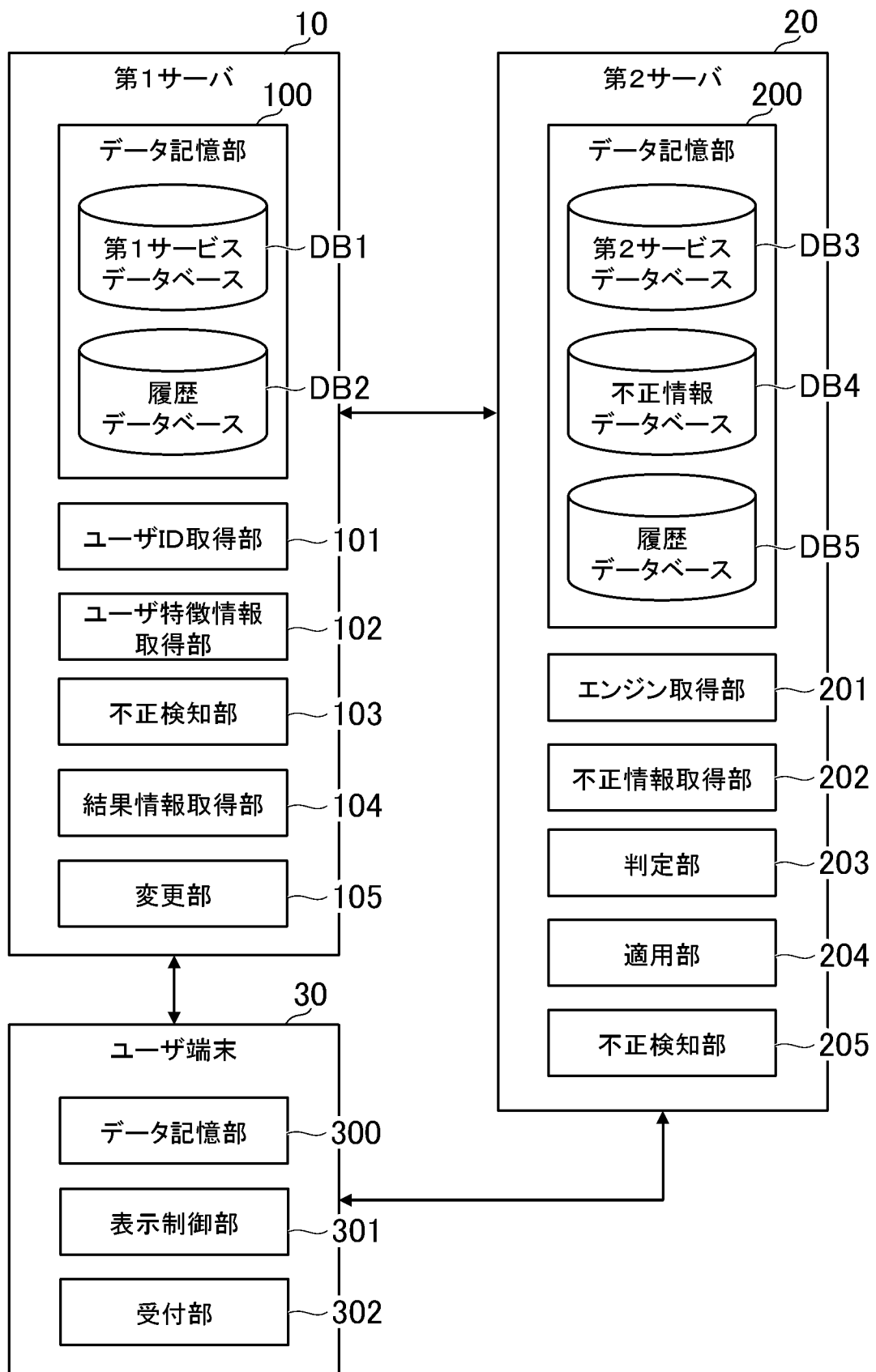
DB5

ユーザ特徴情報	結果情報	ステータス情報
ユーザ特徴情報2-1	正当	確定
ユーザ特徴情報2-2	不正	確定
ユーザ特徴情報2-3	正当	未確定
.	.	.
.	.	.
.	.	.

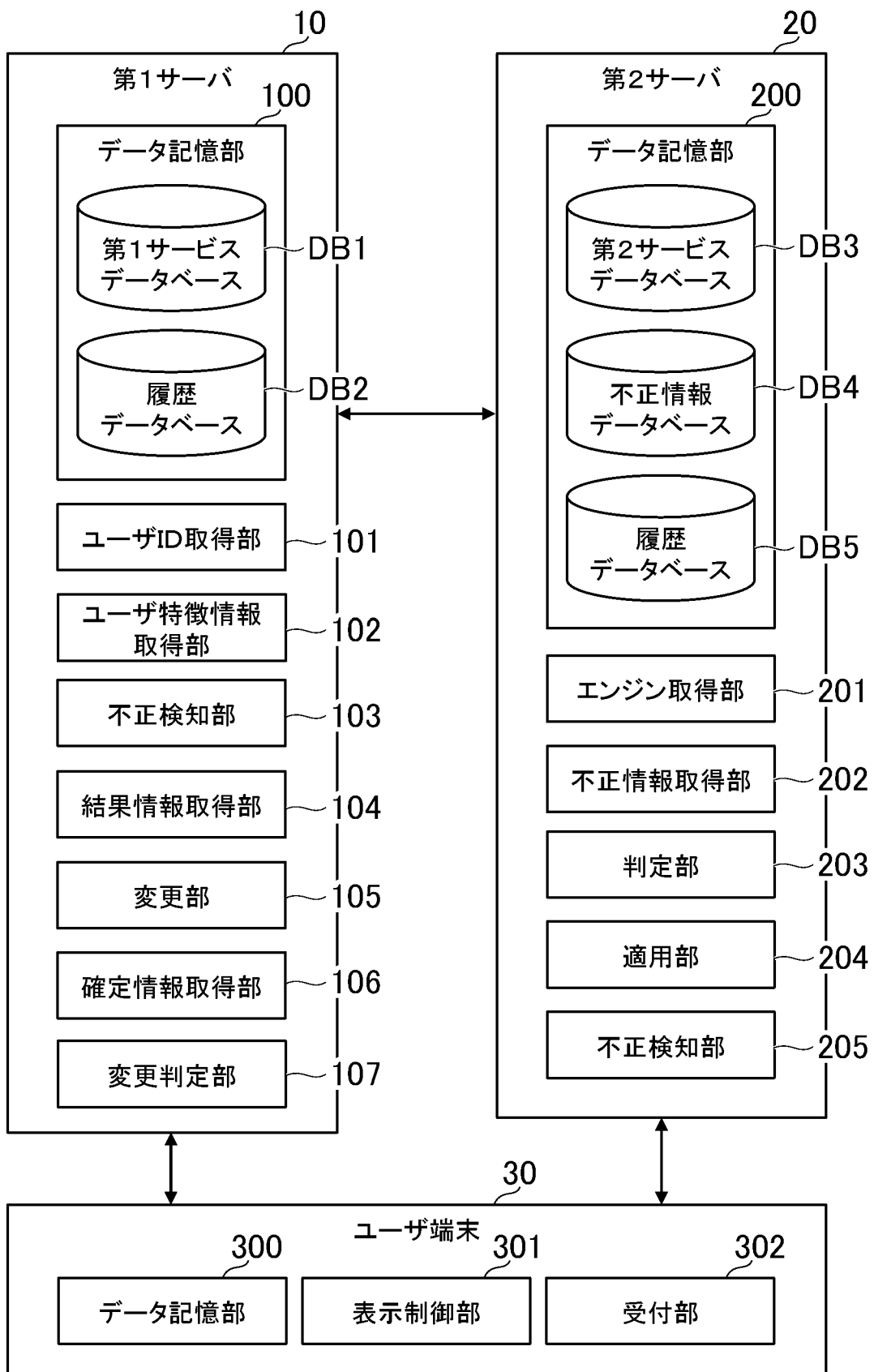
[図14]



[図15]



[図16]



INTERNATIONAL SEARCH REPORT

International application No.

PCT/JP2021/031999

A. CLASSIFICATION OF SUBJECT MATTER		
<i>G06F 21/55</i> (2013.01)j FI: G06F21/55		
According to International Patent Classification (IPC) or to both national classification and IPC		
B. FIELDS SEARCHED		
Minimum documentation searched (classification system followed by classification symbols) G06F21/55		
Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched Published examined utility model applications of Japan 1922-1996 Published unexamined utility model applications of Japan 1971-2021 Registered utility model specifications of Japan 1996-2021 Published registered utility model applications of Japan 1994-2021		
Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)		
C. DOCUMENTS CONSIDERED TO BE RELEVANT		
Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	WO 2020/261426 A1 (RAKUTEN INC) 30 December 2020 (2020-12-30) paragraphs [0095], [0099], [0103]	1-10, 12-15
A		11
<input type="checkbox"/> Further documents are listed in the continuation of Box C. <input checked="" type="checkbox"/> See patent family annex.		
* Special categories of cited documents: "A" document defining the general state of the art which is not considered to be of particular relevance "E" earlier application or patent but published on or after the international filing date "L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified) "O" document referring to an oral disclosure, use, exhibition or other means "P" document published prior to the international filing date but later than the priority date claimed "T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention "X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone "Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art "&" document member of the same patent family		
Date of the actual completion of the international search 02 November 2021		Date of mailing of the international search report 16 November 2021
Name and mailing address of the ISA/JP Japan Patent Office (ISA/JP) 3-4-3 Kasumigaseki, Chiyoda-ku, Tokyo 100-8915 Japan		Authorized officer Telephone No.

INTERNATIONAL SEARCH REPORT
Information on patent family members

International application No.

PCT/JP2021/031999

Patent document cited in search report	Publication date (day/month/year)	Patent family member(s)	Publication date (day/month/year)
WO 2020/261426 A1	30 December 2020	US 2021/0264299 A1 paragraphs [0116], [0121], [0126] TW 202105303 A	

A. 発明の属する分野の分類（国際特許分類（IPC）） G06F 21/55(2013.01)i FI: G06F21/55		
B. 調査を行った分野 調査を行った最小限資料（国際特許分類（IPC）） G06F21/55 最小限資料以外の資料で調査を行った分野に含まれるもの 日本国実用新案公報 1922 - 1996年 日本国公開実用新案公報 1971 - 2021年 日本国実用新案登録公報 1996 - 2021年 日本国登録実用新案公報 1994 - 2021年		
国際調査で使用した電子データベース（データベースの名称、調査に使用した用語）		
C. 関連すると認められる文献		
引用文献の カテゴリー*	引用文献名 及び一部の箇所が関連するときは、その関連する箇所の表示	関連する 請求項の番号
X	WO 2020/261426 A1 (楽天株式会社) 30.12.2020 (2020 - 12 - 30) 段落[0095], [0099], [0103]	1-10, 12-15
A		11
<input type="checkbox"/> C欄の続きにも文献が列挙されている。 <input checked="" type="checkbox"/> パテントファミリーに関する別紙を参照。		
* 引用文献のカテゴリー “A” 特に関連のある文献ではなく、一般的技術水準を示すもの “E” 国際出願日前の出願または特許であるが、国際出願日以後に公表されたもの “L” 優先権主張に疑義を提起する文献又は他の文献の発行日若しくは他の特別な理由を確立するために引用する文献（理由を付す） “O” 口頭による開示、使用、展示等に言及する文献 “P” 国際出願日前で、かつ優先権の主張の基礎となる出願の日の後に公表された文献	“T” 国際出願日又は優先日後に公表された文献であって出願と抵触するものではなく、発明の原理又は理論の理解のために引用するもの “X” 特に関連のある文献であって、当該文献のみで発明の新規性又は進歩性がないと考えられるもの “Y” 特に関連のある文献であって、当該文献と他の1以上の文献との、当業者にとって自明である組合せによって進歩性がないと考えられるもの “&” 同一パテントファミリー文献	
国際調査を完了した日 02.11.2021	国際調査報告の発送日 16.11.2021	
名称及びあて先 日本国特許庁(ISA/JP) 〒100-8915 日本国 東京都千代田区霞が関三丁目4番3号	権限のある職員（特許庁審査官） 平井 誠 5S 9071 電話番号 03-3581-1101 内線 3546	

国際調査報告
パテントファミリーに関する情報

国際出願番号

PCT/JP2021/031999

引用文献	公表日	パテントファミリー文献	公表日
WO 2020/261426 A1	30.12.2020	US 2021/0264299 A1 pars. [0116], [0121], [0126] TW 202105303 A	