(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
2018年5月31日(31.05.2018)

WIPO | PCT

(10) 国際公開番号
WO 2018/096685 A1

(51) 国際特許分類:
H04L 12/66 (2006.01)

(21) 国際出願番号： PCT/JP2016/085227

(22) 国際出願日： 2016年11月28日(28.11.2016)

(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: 株式会社ＰＦＵ (**PFU LIMITED**) [JP/JP]; 〒9291192 石川県かほく市宇野気ヌ９８番地の２ Ishikawa (JP).

(72) 発明者: 広橋　浩司 (**HIROHASHI, Koji**); 〒9291192 石川県かほく市宇野気ヌ９８番地の２ 株式会社ＰＦＵ内 Ishikawa (JP).

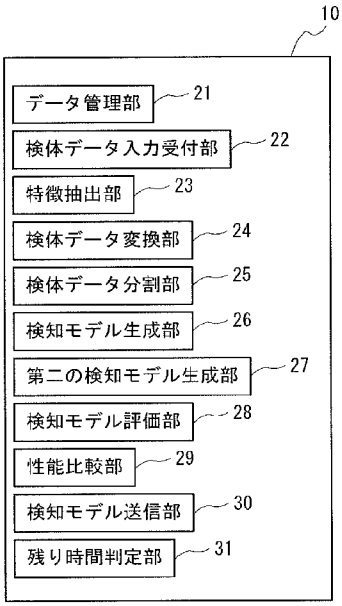
(74) 代理人: 平川 明，外 (**HIRAKAWA, Akira** et al.); 〒1030004 東京都中央区東日本橋３丁目４番１０号 アクロポリス２１ビル８階 Tokyo (JP).

(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, JP, KE, KG, 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, SM, ST, SV,

WO 2018/096685 A1

(54) **Title:** INFORMATION PROCESSING DEVICE, METHOD, AND PROGRAM

(54) 発明の名称： 情報処理装置、方法およびプログラム



21 Data management unit
22 Sample data input reception unit
23 Feature extraction unit
24 Sample data conversion unit
25 Sample data division unit
26 Detection model generation unit
27 Second detection model generation unit
28 Detection model evaluation unit
29 Performance comparison unit
30 Detection model transmission unit
31 Remaining time determination unit

(57) **Abstract:** The present invention addresses the problem of efficiently generating a high-performance detection model by using limited sample data. The information processing device according to the present invention includes: a sample data input reception unit for receiving an input of new sample data relating to a predetermined type of communication; a sample data division unit for dividing the new sample data into mechanical learning sample data and test sample data by using a plurality of mutually different division patterns; a detection model generation unit for generating, for each

[続葉有]

SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, 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))

division pattern, a detection model for detecting a predetermined type of communication by mechanical learning using the mechanical learning sample data; and a detection model evaluation unit for evaluating performance of the detection model by using the test sample data. The detection model generation unit sets the highest-performance model, among the generated detection models, as a new detection model.

(57) 要約:限られた検体データを用いて、性能の高い検知モデルを効率的に生成することを課題とする。情報処理装置に、所定の種類の通信に係る新たな検体データの入力を受け付ける検体データ入力受付部と、新たな検体データを、互いに異なる複数の分割パターンを用いて機械学習用検体データとテスト用検体データに分割する検体データ分割部と、機械学習用検体データを用いた機械学習によって、分割パターン毎に、所定の種類の通信を検知するための検知モデルを生成する検知モデル生成部と、テスト用検体データを用いて検知モデルの性能を評価する検知モデル評価部と、を備え、検知モデル生成部は、生成された検知モデルのうち最も性能の高いモデルを新検知モデルとする。

明　細　書

発明の名称 ： 情報処理装置、方法およびプログラム

技術分野

[0001] 本開示は、所定の種類の通信を検知するための技術に関する。

背景技術

[0002] 従来、ネットワーク上を伝送するトラヒックデータを入力し、学習データ作成装置により作成された学習データと前記トラヒックデータから変数を選択する変数選択手段と、該選択された変数をニューラルネットや決定木等の解析アルゴリズムを用いて解析を行い、パターンを生成する処理手段と、該生成されたパターンを用いて前記解析結果を評価する評価手段とを有し、前記変数選択手段および処理手段、評価手段における処理を１回以上行うことにより、侵入の検知に有効なパターンを生成して異常なトラヒックデータを検知する侵入検知システムが提案されている（特許文献１を参照）。

[0003] また、その他にも、サイバー攻撃を検知するために、機械学習を用いるシステムが種々提案されている（特許文献２から４を参照）。

先行技術文献

特許文献

[0004] 特許文献1：特開２００４－３１２０８３号公報

特許文献2：特開２０１６－９１５４９号公報

特許文献3：米国特許出願公開第２０１５／０１２８２６３号明細書

特許文献4：米国特許出願公開第２００４／００５９９４７号明細書

発明の概要

発明が解決しようとする課題

[0005] 従来、サイバー攻撃（マルウェアを含む）等の所定の種類の通信を検知するために、検知対象とパターンファイルをマッチングさせるパターンマッチング検知や、プログラムやソフトウェアの挙動に基づいて判定する振る舞い検知などが実施されている。このような検知技術を実施するにあたっては、

日々進化を続け種類が増え続けている通信の種類に対応するため、検知機能の継続的な更新が重要である。

[0006] しかし、検知機能の更新に伴う作業には、検知ロジックの更新、および更新後の検知ロジックの評価等の、人が介在すると共に時間がかかる作業が必要である。例えば、検知ロジックの更新では、新たな種類の通信の検知に対応するために、人が、実際の通信のデータから法則を見出した上で新たな検知ロジックを設計し、それを基にコーディングを実施する作業が行われる。また、更新後の検知ロジックの評価では、更新後の検知ロジックが、新たな種類の通信を検知できるかどうかを、人が評価する作業が行われる。このため、検知機能の更新を継続的に実施するには、人の負担が大きい。

[0007] ここで、検知ロジックを機械学習によって自動生成された検知モデルに置き換えることで、上記の作業を自動化することが考えられる。しかし、この場合、量が限られた検体データを、検知モデルの生成に使用する機械学習用検体データと、検知モデルの性能（検知精度）の比較に使用するテスト用検体データとに分ける必要がある。そして、新しく入手した検体データを機械学習用とテスト用に分ける場合、どのように分けるのが、性能の高い検知モデルを生成するために最適であるかについては、検体データの種類や数などを含む環境によって異なる。

[0008] 本開示は、上記した問題に鑑み、限られた検体データを用いて、性能の高い検知モデルを効率的に生成することを課題とする。

課題を解決するための手段

[0009] 本開示の一例は、所定の種類の通信に係る新たな検体データの入力を受け付ける検体データ入力受付手段と、前記新たな検体データを、互いに異なる複数の分割パターンを用いて機械学習用検体データとテスト用検体データに分割する検体データ分割手段と、前記機械学習用検体データを用いた機械学習によって、分割パターン毎に、前記所定の種類の通信を検知するための検知モデルを生成する検知モデル生成手段と、前記テスト用検体データを用いて前記検知モデルの性能を評価する検知モデル評価手段と、を備え、前記検

知モデル生成手段は、生成された検知モデルのうち最も性能の高いモデルを新検知モデルとする、情報処理装置である。

[0010] 本開示は、情報処理装置、システム、コンピューターによって実行される方法またはコンピューターに実行させるプログラムとして把握することが可能である。また、本開示は、そのようなプログラムをコンピューターその他の装置、機械等が読み取り可能な記録媒体に記録したものとしても把握できる。ここで、コンピューター等が読み取り可能な記録媒体とは、データやプログラム等の情報を電気的、磁気的、光学的、機械的または化学的作用によって蓄積し、コンピューター等から読み取ることができる記録媒体をいう。

発明の効果

[0011] 本開示によれば、限られた検体データを用いて、性能の高い検知モデルを効率的に生成することが可能となる。

図面の簡単な説明

[0012] [図1]実施形態に係るシステムの構成を示す概略図である。

[図2]実施形態に係る検知モデル生成装置およびネットワーク監視装置のハードウェア構成を示す図である。

[図3]実施形態に係る検知モデル生成装置の機能構成の概略を示す図である。

[図4]実施形態に係る更新用検知モデル生成処理の流れの概要を示すフローチャートである。

[図5]実施形態に係る特徴抽出処理の流れの概要を示すフローチャートである。

[図6]実施形態に係る検体データ分割処理（メイン処理）の流れの概要を示すフローチャートである。

[図7]実施形態に係る検体データ分割処理（サブ処理Ａ）の流れの概要を示すフローチャートである。

[図8]実施形態に係る検体データ分割処理（サブ処理Ｂ）の流れの概要を示すフローチャートである。

[図9]実施形態に係る第二の新検知モデル生成処理の流れの概要を示すフロー

チャートである。

[図10]実施形態に係る性能比較処理（メイン処理）の流れの概要を示すフローチャートである。

[図11]実施形態に係る性能比較処理（サブ処理）の流れの概要を示すフローチャートである。

発明を実施するための形態

[0013] 以下、本開示に係る情報処理装置、方法およびプログラムの実施の形態を、図面に基づいて説明する。但し、以下に説明する実施の形態は、実施形態を例示するものであって、本開示に係る情報処理装置、方法およびプログラムを以下に説明する具体的構成に限定するものではない。実施にあたっては、実施の態様に応じた具体的構成が適宜採用され、また、種々の改良や変形が行われてよい。

[0014] 本実施形態では、本開示に係る情報処理装置、方法およびプログラムを、ネットワークを監視し、主にサイバー攻撃に係る悪意の通信を所定の種類の通信として検知するためのシステムにおいて実施した場合の実施の形態について説明する。但し、本開示に係る情報処理装置、方法およびプログラムは、所定の種類の通信を検知するための技術について広く用いることが可能であり、本開示の適用対象は、本実施形態において示した例に限定されない。例えば、検知の対象は、所定の種類の通信であればよく、正常な通信が検知の対象とされてもよい。

[0015] ＜システムの構成＞

図１は、本実施形態に係るシステム１の構成を示す概略図である。本実施形態に係るシステム１は、所定の種類の通信の検知に用いられる検知モデルを自動的に生成する検知モデル生成装置１０と、検知モデルを用いて所定の種類の通信を検知する複数のネットワーク監視装置２０Ａおよび２０Ｂと、を備える。なお、何れのネットワーク監視装置であるかを特定しない場合には単に「ネットワーク監視装置２０」と記載する。図１にはネットワーク監視装置２０を２つのみ図示しているが、３以上であってもよい。また、複数

のネットワーク監視装置２０Ａおよび２０Ｂの夫々は、複数の情報処理端末９０（以下、「ノード９０」と称する）が接続されるネットワークセグメント２Ａおよび２Ｂ（何れのネットワークセグメントであるかを特定しない場合には単に「ネットワークセグメント２」と記載する。）に配置され、各ネットワークセグメント２においてノード９０に係る通信を監視する。

[0016] ネットワークセグメント２内のノード９０およびネットワーク監視装置２０は、インターネットや広域ネットワークを介して遠隔地において接続された各種のサーバー（検知モデル生成装置１０を含む）と通信可能である。本実施形態において、ネットワーク監視装置２０は、標的型サイバー攻撃検知技術（例えば、株式会社ＰＦＵの「Ｍａｌｉｃｉｏｕｓ Ｉｎｔｒｕｓｉｏｎ Ｐｒｏｃｅｓｓ Ｓｃａｎ」技術）を搭載したセンサー等の検知装置であり、ネットワークセグメント２のスイッチまたはルータと、その上位にある他のスイッチまたはルータと、の間に接続されることで、通過するパケットやフレーム等を取得し、所定の種類の通信を検知する。この場合、ネットワーク監視装置２０は、取得したパケットのうち、遮断しなくてもよいパケットについては転送するインラインモードで動作する。

[0017] 但し、上記ネットワーク構成は、本開示を実施するための一例であり、実施にあたってはその他のネットワーク構成が採用されてもよい。例えば、ネットワーク監視装置２０は、スイッチまたはルータのモニタリングポート（ミラーポート）に接続されることで、ノード９０によって送受信されるパケットやフレーム等を取得してもよい（図示は省略する）。この場合、ネットワーク監視装置２０は、取得したパケットを転送しないパッシブモードで動作する。また、例えば、ネットワーク監視装置２０は、モニタリングポート（ミラーポート）に接続されず、単にネットワークセグメント２に接続されている場合であっても、ネットワークセグメント２を流れるフレームを、自身のＭＡＣアドレス宛でないものも含めて全て取得することで、ノード９０によって送受信されるパケットやフレーム等を取得することが出来る。この場合も、ネットワーク監視装置２０は、パッシブモードで動作する。また、

例えば、ネットワーク監視装置２０は、ルータまたはスイッチに内包されてもよい。

[0018] 図２は、本実施形態に係る検知モデル生成装置１０およびネットワーク監視装置２０のハードウェア構成を示す図である。検知モデル生成装置１０は、ＣＰＵ（Ｃｅｎｔｒａｌ Ｐｒｏｃｅｓｓｉｎｇ Ｕｎｉｔ）１１、ＲＯＭ（Ｒｅａｄ Ｏｎｌｙ Ｍｅｍｏｒｙ）１２、ＲＡＭ（Ｒａｎｄｏｍ Ａｃｃｅｓｓ Ｍｅｍｏｒｙ）１３、ＥＥＰＲＯＭ（Ｅｌｅｃｔｒｉｃａｌｌｙ Ｅｒａｓａｂｌｅ ａｎｄ Ｐｒｏｇｒａｍｍａｂｌｅ Ｒｅａｄ Ｏｎｌｙ Ｍｅｍｏｒｙ）やＨＤＤ（Ｈａｒｄ Ｄｉｓｋ Ｄｒｉｖｅ）等のストレージ１４、ＮＩＣ（Ｎｅｔｗｏｒｋ Ｉｎｔｅｒｆａｃｅ Ｃａｒｄ）１５等の通信ユニット、等を備えるコンピューターである。

[0019] また、ネットワーク監視装置２０は、ＣＰＵ、ＲＯＭ、ＲＡＭ、ストレージ、通信ユニット、等を備えるコンピューターである。ネットワーク監視装置２０は、ネットワークに接続された端末によって送受信される通信のデータを取得し、プロトコルに応じて定められた検査項目に従ってデータを検査する。例えば、ネットワーク監視装置２０は、フローに係るプロトコルがＨＴＴＰであると判定された場合に、ＨＴＴＰについて定められた検査項目に従ってデータを検査する。ネットワーク監視装置２０は、この検査の際に、検知モデル生成装置１０によって生成・配信された検知モデルを用いる。

[0020] そして、ネットワーク監視装置２０は、サイバー攻撃等のイベントが検知された場合に、当該イベントに係るデータ（証跡）を保存または当該イベントの検知をユーザーに通知する。なお、本実施形態において、ネットワーク監視装置２０による監視および検知の対象となる「端末」には、ネットワークセグメント２に接続されたノード９０の他、ノード９０とルータを介して通信するその他の装置（他のネットワークに属するノードや外部サーバー等）が含まれる。

[0021] 図３は、本実施形態に係る検知モデル生成装置１０の機能構成の概略を示す図である。なお、図３においては、検知モデル生成装置１０以外の構成に

ついては、図示を省略している。検知モデル生成装置１０は、ストレージ１４に記録されているプログラムが、ＲＡＭ１３に読み出され、ＣＰＵ１１によって実行されることで、データ管理部２１、検体データ入力受付部２２、特徴抽出部２３、検体データ変換部２４、検体データ分割部２５、検知モデル生成部２６、第二の検知モデル生成部２７、検知モデル評価部２８、性能比較部２９、検知モデル送信部３０および残り時間判定部３１を備える情報処理装置として機能する。なお、本実施形態では、検知モデル生成装置１０の備える各機能は、汎用プロセッサであるＣＰＵ１１によって実行されるが、これらの機能の一部または全部は、１または複数の専用プロセッサによって実行されてもよい。また、これらの機能の一部または全部は、クラウド技術等を用いて、遠隔地に設置された装置や、分散設置された複数の装置によって実行されてもよい。

[0022] データ管理部２１は、新検知モデルの生成後に、新たな検体データおよび新規の特徴を、所定の環境に関連づけて保存することで、当該所定の環境において所定の種類の通信を検知するための現検知モデル（ネットワーク監視装置２０が現在使用している検知モデル）の生成に用いられた既存の特徴および既存の検体データを蓄積する。本実施形態では、既存の特徴は特徴データ管理データベースに、既存の検体データは検体データ管理データベースに、蓄積される。なお、特徴データ管理データベースには、何れの環境でも用いることを示す環境識別子「共通」が付された特徴データが、予め蓄積されており、検体データ管理データベースには、何れの環境でも用いることを示す環境識別子「共通」が付された検体データが、予め蓄積されている。

[0023] 検体データ入力受付部２２は、所定の種類の通信に係る新たな検体データの入力を受け付ける。

[0024] 特徴抽出部２３は、新たな検体データから、既存の特徴とは異なる新規の特徴を抽出する。

[0025] 検体データ変換部２４は、既存の特徴および新規の特徴を使用して、既存の検体データおよび新たな検体データを、機械学習のための所定のデータ形

式に変換する。この際、検体データ変換部２４は、検知モデル生成部２６によって用いられる機械学習の方式に適したデータ形式となるように、検体データを変換する。例えば、検知モデル生成部２６がランダムフォレスト方式の機械学習によって検知モデルを生成する場合、検体データ変換部２４は、Ｂａｇ－ｏｆ－Ｗｏｒｄｓ（ＢｏＷ）方式で検体データを変換し、検知モデル生成部２６がニューラルネットワーク方式の機械学習（例えば、畳み込みニューラルネットワークやリカレントニューラルネットワーク）によって検知モデルを生成する場合、検体データ変換部２４は、検体データ内のデータを各特徴に対応する一意の数値に変換する方法で検体データを変換する。

[0026] 例えば、変換方式がＢｏＷであり、検体データがＨＴＴＰリクエストヘッダー群である場合、検体データ変換部２４は、ＨＴＴＰリクエストヘッダー単位で、既存および新規の各特徴の出現回数を求め、それらを次元として持つベクトルデータに変換する（ＢｏＷベクトルの生成）。より具体的には、特徴が「Ｈｏｓｔ：」「ａｂｃ」「ｅｆｇ」である場合、変換前データ「Ｈｏｓｔ：　ａｂｃ」はＢｏＷベクトル［１，　１，　０］に変換され、変換前データ「Ｈｏｓｔ：　ｅｆｇ」はＢｏＷベクトル［１，　０，　１］に変換される。

[0027] 検体データ分割部２５は、新たな検体データを、互いに異なる複数の分割パターンを用いて機械学習用検体データとテスト用検体データに分割する。本実施形態では、検体データの分割処理を継続するか否かを判定するために、検知モデル毎に算出される性能指標（以下、実施例中では「Ａｃｃｕｒａｃｙ」と称する）が用いられる。本実施形態では、図６を用いて後述する判定条件（｜Ｍａ－Ｍａ´｜＞０．０００１）の通り、検体データ分割部２５は、「第一の分割パターンを用いて生成された検知モデルの性能」と、「当該第一の分割パターンより機械学習用検体データを増やした第二の分割パターンを用いて生成された検知モデルの性能」との間に所定以上の差がある場合、当該第二の分割パターンよりも機械学習用検体データを増やした第三の分割パターンを更に生成する。

[0028] 但し、処理続行可否の判定には、その他の判定条件が用いられてもよい。例えば、「｜Ｍａ－Ｍａ´｜＞０．０００１かつＭａ´＜Ｍａ」の判定条件では、検体データ分割部２５は、「第一の分割パターンを用いて生成された検知モデルの性能」に比べて、「当該第一の分割パターンより機械学習用検体データを増やした第二の分割パターンを用いて生成された検知モデルの性能」が所定以上優れている場合、当該第二の分割パターンよりも機械学習用検体データを増やした第三の分割パターンを更に生成する。この判定条件（｜Ｍａ－Ｍａ´｜＞０．０００１かつＭａ´＜Ｍａ）を用いた場合、分割数を上げて性能が落ちた時点で分割処理が終了するため、図６を用いて後述する判定条件（｜Ｍａ－Ｍａ´｜＞０．０００１）に比べて早期に新検知モデルを決定することができる。

[0029] 検知モデル生成部２６は、機械学習用検体データおよび既存の検体データを用いた機械学習によって、分割パターン毎に、所定の種類の通信を検知するための検知モデルを生成し、生成された検知モデルのうち最も性能の高いモデルを新検知モデルとする。ここで、機械学習では、所定のデータ形式に変換された検体データが用いられる。

[0030] 第二の検知モデル生成部２７は、検知モデル生成部２６とは異なる機械学習アルゴリズムまたは異なる機械学習用パラメータを用いて第二の新検知モデルを生成する。ここで、検知モデル生成装置１０は、複数の第二の検知モデル生成部２７．２から２７．ｍを備えてよい。複数の第二の検知モデル生成部２７．２から２７．ｍは、各々が互いに事なる機械学習アルゴリズムまたは機械学習用パラメータを用いることで、互いに異なる第二の新検知モデルを生成する。

[0031] 検知モデル評価部２８は、テスト用検体データを用いて検知モデルの性能を評価する。

[0032] 性能比較部２９は、所定の種類の通信を検知するための現検知モデルと１または複数の新検知モデルの性能とを比較し、現検知モデルと概略同等（例えば、誤差の範囲内）かまたは現検知モデルよりも大きい性能を有する新検

知モデルを、現検知モデルを置き換えるための更新用検知モデルに決定する。

[0033] 検知モデル送信部３０は、更新用検知モデルをネットワーク監視装置２０に送信し、当該ネットワーク監視装置２０において用いられる検知モデルを更新させる。

[0034] 残り時間判定部３１は、検体データ分割部２５が異なる分割パターンを用いて新たな検体データを分割する毎に、更新用検知モデルが得られるまでの予め設定された上限時間までの残り時間（Ｔｒ）と処理見積時間（Ｔｒ´－Ｔｒ）を計算し、当該処理見積時間が残り時間を超えるか否かを判定する。本実施形態において、処理見積時間が残り時間を超える場合、検体データ分割部２５は、処理を終了する。

[0035] なお、本実施形態では、処理見積時間が残り時間を超える場合、分割処理自体を失敗とみなして終了することとしているが、検体データ分割部２５および検知モデル生成部２６は、処理見積時間が残り時間を超える場合、ここまでの処理で生成された検知モデルを用いて新検知モデルを決定してもよい。

[0036] ＜処理の流れ＞

次に、本実施形態に係るシステム１によって実行される処理の流れを、フローチャートを用いて説明する。なお、以下に説明するフローチャートに示された処理の具体的な内容および処理順序は、本開示を実施するための一例である。具体的な処理内容および処理順序は、本開示の実施の形態に応じて適宜選択されてよい。

[0037] 図４は、本実施形態に係る更新用検知モデル生成処理の流れの概要を示すフローチャートである。本実施形態に係る更新用検知モデル生成処理は、所定の環境のサイバー攻撃の検体データが検知モデル生成装置１０に入力されたことを契機として実行される。

[0038] ステップＳ００１では、サイバー攻撃の検体データの入力が受け付けられる。検体データ入力受付部２２は、所定の種類の通信（ここでは、サイバー

攻撃）を検知するための現検知モデルが用いられている所定の環境（例えば、ネットワーク監視装置２０Ａが設置されたローカルネットワーク）のサイバー攻撃の検体データの入力を受け付ける。検体データ入力受付部２２は、入力されたサイバー攻撃の検体データを、新たな検体データとして、ＲＡＭ１３またはストレージ１４に記録する。なお、本実施形態では、ネットワーク監視装置２０Ａが設置されたローカルネットワークを「環境Ａ」と称し、ネットワーク監視装置２０Ａで用いられる更新用検知モデルの生成について説明する。なお、ここで受け付けられる検体データは、サイバー攻撃の有無について分類済みのデータであり、検体データには、正解ラベル（サイバー攻撃の有無を示す情報）が付されている。その後、処理はステップＳ００２へ進む。

[0039] ステップＳ００２では、特徴抽出処理が実行される。特徴抽出部２３は、入力された新たな検体データから、新規の特徴を抽出する。特徴抽出処理の詳細については、図５を用いて後述する。その後、処理はステップＳ００３へ進む。

[0040] ステップＳ００３およびステップＳ００４では、検体データ分割処理が実行され、分割の成否が判定される。検体データ分割部２５は、新たな検体データを、機械学習用とテスト用とに分割し、また、検知モデル生成部２６は、分割の過程で、新検知モデルを１つ生成する（Ｓ００３）。即ち、本実施形態では、検体データをどのように機械学習用とテスト用とに分割するかが決定した時点で、少なくとも１つの新検知モデルが生成される。検体データ分割処理の詳細については、図６から図８を用いて後述する。ステップＳ００４では、ステップＳ００１で入力された新たな検体データの分割の成否が判定される。新たな検体データの分割に失敗した場合、本フローチャートに示された処理は終了する。一方、新たな検体データの分割に成功した場合、処理はステップＳ００５およびステップＳ００６．２からＳ００６．ｍへ進む。

[0041] ステップＳ００５では、既存の特徴を使用した検体データの変換が行われ

る。検体データ変換部２４は、既存の特徴を使用して、全検体データ（既存の検体データ＋新たな検体データ）を、機械学習アルゴリズムの入力データに適した形式に変換する。その後、処理はステップＳ００７へ進む。

[0042] ステップＳ００６．２からＳ００６．ｍでは、既存の特徴および新規の特徴を使用した検体データの変換、および第二の新検知モデルの生成が行われる。なお、本実施形態において、ステップＳ００６．２からＳ００６．ｍの各処理は、単一のサーバーまたは複数のサーバーで並列に処理される。各サーバーで処理される第二の新検知モデル生成処理の詳細については、図９を用いて後述する。その後、処理はステップＳ００７へ進む。

[0043] ステップＳ００７およびステップＳ００８では、性能比較処理が実行され、現検知モデルの置き換え対象となる新検知モデル（更新用検知モデル）が存在するか否かが判定される。性能比較部２９は、これまでに生成された１または複数の新検知モデル（第二の新検知モデルを含む）と現検知モデルとを比較し、比較の結果に基づいて、現検知モデルを置き換えるための更新用検知モデルを決定する（ステップＳ００７）。性能比較処理の詳細については、図１０および図１１を用いて後述する。性能比較処理の結果、現検知モデルの置き換え対象となる新検知モデル（更新用検知モデル）が存在しない場合、本フローチャートに示された処理は終了する。一方、性能比較処理の結果、現検知モデルの置き換え対象となる新検知モデル（更新用検知モデル）が存在する場合、処理はステップＳ００９へ進む。

[0044] ステップＳ００９では、データ管理処理が実行される。データ管理部２１は、新たな検体データおよびその正解ラベル（サイバー攻撃の有無を示す情報）を、互いに関連づけて、環境識別子「環境Ａ」のデータとして検体データ管理データベースに保存する。また、データ管理部２１は、新規の特徴を、環境識別子「環境Ａ」のデータとして特徴データ管理データベースに保存する。その後、処理はステップＳ０１０へ進む。

[0045] ステップＳ０１０では、更新用検知モデルがネットワーク監視装置２０へ送信される。検知モデル送信部３０は、現検知モデルの置き換え対象となる

新検知モデル（更新用検知モデル）を、ネットワーク監視装置２０へ送信する。そして、更新用検知モデルを受信したネットワーク監視装置２０は、受信した更新用検知モデルを新たな現検知モデルとし、以降、新たな現検知モデルを用いて、サイバー攻撃を検知する。その後、本フローチャートに示された処理は終了する。

[0046] 上記説明したような更新用検知モデル生成処理が実行されることで、限られた検体データを用いて、性能の高い検知モデルを効率的に生成することが可能となる。以下、上記フローチャートに登場する各ステップの詳細について、フローチャートを参照しながら説明する。

[0047] 図５は、本実施形態に係る特徴抽出処理の流れの概要を示すフローチャートである。本フローチャートは、更新用検知モデル生成処理（図４を参照）のステップＳ００２の処理を詳細に説明するためのものである。

[0048] ステップＳ１０１では、既存の検体データが取り出される。特徴抽出部２３は、環境識別子が「共通」に該当する検体データおよび「環境Ａ」に該当する検体データと正解ラベル（サイバー攻撃の有無を示す情報）を、検体データ管理データベースから取り出す。その後、処理はステップＳ１０２へ進む。

[0049] ステップＳ１０２では、既存の特徴が取り出される。特徴抽出部２３は、環境識別子が「共通」に該当する検体データおよび「環境Ａ」に該当する特徴データを、特徴データ管理データベースから取り出す。その後、処理はステップＳ１０３へ進む。

[0050] ステップＳ１０３では、新たな検体データの分解処理が行われる。特徴抽出部２３は、分解ルールとして正規表現を利用するなどの方法で、新たな検体データの内容を単語毎に分解する。例えば、検体データがＨＴＴＰリクエストヘッダー群である場合、特徴抽出部２３は、以下の正規表現を分解ルールとして用いて、リクエストヘッダー群の内容を、単語毎に分解することが出来る。

[a-zA-Z0-9:][a-zA-Z0-9_¥¥-:.]+¥¥b

その後、処理はステップＳ１０４へ進む。

[0051] ステップＳ１０４では、新規の特徴が抽出される。特徴抽出部２３は、新たな検体データと既存の特徴とを突き合わせることで、新たな検体データから新規の特徴（既存の特徴に存在しない特徴）を抽出する。その後、本フローチャートに示された処理は終了する。

[0052] 図６は、本実施形態に係る検体データ分割処理（メイン処理）の流れの概要を示すフローチャートである。本フローチャートは、更新用検知モデル生成処理（図４を参照）のステップＳ００３の処理を詳細に説明するためのものである。

[0053] ステップＳ２０１では、既存の特徴および新規の特徴を使用した検体データの変換が行われる。検体データ変換部２４は、既存の特徴および新規の特徴を使用して、全検体データ（既存の検体データ＋新たな検体データ）を、機械学習アルゴリズムの入力データに適した形式に変換する。その後、処理はステップＳ２０２へ進む。

[0054] ステップＳ２０２では、変数が初期化される。検体データ分割部２５は、処理に用いる各変数に初期値を設定する。具体的には、検体データ分割部２５は、後述するステップＳ２０３からステップＳ２１２のループ処理における１ループ前の分割数の時に算出された平均Ａｃｃｕｒａｃｙ（Ｍａ´）に０を、１ループ前の分割数の時に見積もられた残り時間（Ｔｒ´）に０を、分割数（ｎ）に２を設定する。その後、処理はステップＳ２０３へ進む。

[0055] ステップＳ２０３では、検体データが分割される。検体データ分割部２５は、新たな検体データをｎ個に分割する（検体データ分割処理（サブ処理Ａ））。検体データ分割処理（サブ処理Ａ）の詳細については、図７を用いて後述する。その後、処理はステップＳ２０４へ進む。

[0056] ステップＳ２０４およびステップＳ２０５では、分割が成功した場合に、検知モデルが生成され、平均Ａｃｃｕｒａｃｙ（Ｍａ）が算出される。ステップＳ２０３における検体データ分割処理（サブ処理Ａ）において分割に失敗した場合、本フローチャートに示された処理は終了する（分割失敗）。一

方、分割に成功した場合、検知モデル生成部２６は、ｎ個の検知モデルを生成し、検体データ分割部２５は、平均Ａｃｃｕｒａｃｙ（Ｍａ）を算出する（検体データ分割処理（サブ処理Ｂ））。これは、新たな検体データをｎ分割した場合、Ａｃｃｕｒａｃｙ算出用に機械学習用検体データの分割グループを選択する方法はｎ通りあり、ｎ個の検知モデルを生成できるためである。検知モデル評価部２８は、ｎ個の検知モデルの夫々についてＡｃｃｕｒａｃｙを算出し、検体データ分割部２５は、その平均値を平均Ａｃｃｕｒａｃｙ（Ｍａ）とする。検体データ分割処理（サブ処理Ｂ）の詳細については、図８を用いて後述する。その後、処理はステップＳ２０６へ進む。

[0057] ステップＳ２０６では、分割数ｎを増やしたことで平均Ａｃｃｕｒａｃｙに所定以上の変化が生じたか否かが判定される。検体データ分割部２５は、１ループ前の分割数の時に算出された平均Ａｃｃｕｒａｃｙ（Ｍａ´）と、現在の分割数（ｎ）の時に算出された平均Ａｃｃｕｒａｃｙ（Ｍａ）との間に、所定以上の差（例えば、所定値よりも大きな差）があるか否かを判定する。本実施形態では、例えば、以下の数式を用いて判定を行う。

｜Ｍａ　－　Ｍａ´｜　＞　０．０００１

即ち、この式では、平均Ａｃｃｕｒａｃｙの差が０．０１％以内であるか否かが判定される。この判定に用いられる所定値には、その他の値が採用されてもよい。

[0058] 所定以上の差がある場合、分割数（換言すれば、機械学習用検体データの量）をより増やすことによる平均Ａｃｃｕｒａｃｙの向上可能性が高いと判定し、処理はステップＳ２０７へ進む。一方、所定以上の差がない（同等である）場合、これ以上分割数（換言すれば、機械学習用検体データの量）を増やすことによる平均Ａｃｃｕｒａｃｙの向上可能性は低く、１ループ前の分割数での検体データ分割が適正であると判定し、処理はステップ２１３へ進む。

[0059] 本実施形態では、平均Ａｃｃｕｒａｃｙに所定以上の変化がみられなくなるまで、分割数ｎを増やしながら（機械学習用検体データの量を増やしなが

ら）、ステップＳ２０３からステップＳ２０６の分割処理が繰り返し実行される。これは、同様の特徴を持つ検体データを使用して検知モデルを生成する場合、はじめのうちは、機械学習用検体データの数を増やすことで、最も特徴的な規則性を見つけることにつながり、検知モデルの性能は比較的大きく向上するが、さらに数を増やし、学習効果が少なくなると、検知モデルの性能の増加はなだらか（または一定）になるためである。検体データ分割部２５は、上記判定条件を用ることで、検知モデルの性能の増加がなだらか（または一定）になり始めた時点の学習用検体データの割合（学習用検体データの量）を特定し、これを検知モデルの生成に最低限必要な学習用検体データの割合（学習用検体データの量）とし、残りの検体データをテスト用検体データとする。

[0060] なお、本実施形態では、検体データの分割に、検体データを２のべき乗（２、４、８、１６・・・）個のグループへ均等分割する方式を採用しているが、検体データの分割方式には、その他の方式が用いられてもよい。例えば、検体データ中の、機械学習用検体データとして用いられる検体データの量を一定量や一定割合ずつ増やしていく方式が採用されてもよい。

[0061] ステップＳ２０７からステップＳ２０９では、分割上限数（Ｎｍａｘ）が定義されている場合に、分割数（ｎ）が上限に達したか否かが判定される。検体データ分割部２５は、分割上限数（Ｎｍａｘ）が定義されているか否かを判定する（ステップＳ２０７）。分割上限数（Ｎｍａｘ）が定義されている場合、検体データ分割部２５は、現在の分割数（ｎ）を２倍した値を新たな分割数（ｎ）として設定し、また、今回の処理で算出された平均Ａｃｃｕｒａｃｙ（Ｍａ）の値を、１ループ前の分割数の時に算出された平均Ａｃｃｕｒａｃｙ（Ｍａ´）として設定する（ステップＳ２０８）。そして、新たな分割数（ｎ）が分割上限数（Ｎｍａｘ）を超える場合（ステップＳ２０９のＹＥＳ）、本フローチャートに示された処理は終了する（分割失敗）。一方、新たな分割数（ｎ）が分割上限数（Ｎｍａｘ）を超えない場合（ステップＳ２０９のＮＯ）、処理はステップＳ２０３へ戻り、新たな分割数（ｎ）

を用いた検体データの分割が行われる。

[0062] 分割上限数（Ｎｍａｘ）が未定義である場合、後述する、置き換え候補の検知モデル（更新用検知モデル）を確定させるまでの所要上限時間（Ｔｍａｘ）を超過しない限り、検体データ分割処理は続行される。なお、検知システムの導入環境の実情に柔軟に対応できるよう、分割上限数（Ｎｍａｘ）を任意に定義できる仕組みを備えてもよい。分割上限数（Ｎｍａｘ）が定義された場合、ステップＳ２１０からステップＳ２１２の処理は実行されないため、更新用検知モデルを確定させるまでの時間は所要上限時間（Ｔｍａｘ）に制約されない。

[0063] ステップＳ２１０からステップＳ２１２では、分割上限数（Ｎｍａｘ）が定義されていない場合に、処理に使用できる残り時間に基づいて、処理を継続するか否かが判定される。残り時間判定部３１は、検体データ分割処理で使用できる残り時間（Ｔｒ）を見積もる（ステップＳ２１０）。

[0064] 本実施形態では、以下の数式を用いて、検体データ分割処理をｉ回実行した時点の残り時間（Ｔｒ）を見積もる。

Ｔｒ　＝　Ｔｍａｘ　－　（Ｔｂ　＋　Ｔｓ　＋　Ｔａ）

Ｔａ　＝　ｔ　＊　（ｍ　／　Ｍ）

ここで、（ｍ　／　Ｍ）の値は小数点以下切り上げ

Ｔｍａｘ：更新用検知モデルを確定させるまでの所要上限時間［秒］（特徴抽出処理から性能比較処理までが対象）

Ｔｂ：「検体データ分割処理」の直前までの所要時間［秒］

Ｔｓ：「検体データ分割処理」をｉ回実行した時点の所要時間［秒］

Ｔａ：検知モデルの生成処理および検知性能比較処理の所要見積時間［秒］

ｔ：ｉ回目の検体データ分割処理における「１検知モデルの生成時間＋Ａｃｃｕｒａｃｙ算出時間」の最大値［秒］

ｍ：現検知モデルおよび「第二の新検知モデル生成処理２からｍの処理」で生成する第二の新検知モデルの総数

Ｍ：第二の新検知モデルを並列で生成する際に使用するサーバーの数

なお、所要上限時間（Ｔｍａｘ）は、管理者によって任意に設定可能であってよい。

[0065] 検体データ分割処理で使用できる残り時間（Ｔｒ）が見積もられると、残り時間判定部３１は、１ループ前の分割数の時に見積もられた残り時間（Ｔｒ´）から新たに見積もられた残り時間（Ｔｒ）を減じた値を、処理見積時間（Ｔｒ´－Ｔｒ）とみなして、新たに見積もられた残り時間（Ｔｒ）が、処理見積時間（Ｔｒ´－Ｔｒ）よりも多いか否かを判定する（ステップＳ２１１）。具体的には、以下の等価な条件式の何れかを用いて判定することができる。

Ｔｒ ＞＝ Ｔｒ´ － Ｔｒ

Ｔｒ ＞＝ Ｔｒ´ ／ ２

但し、処理見積時間は、その他の方法で算出されてもよいし、判定条件にはその他の条件式が採用されてもよい。

[0066] 新たに見積もられた残り時間（Ｔｒ）が上記条件を満たさない場合、残り時間（Ｔｒ）の間にもう１回の分割が終わらない可能性が高いため、本フローチャートに示された処理は終了する（分割失敗）。一方、新たに見積もられた残り時間（Ｔｒ）が上記条件を満たす場合、検体データ分割部２５は、現在の分割数（ｎ）を２倍した値を新たな分割数（ｎ）として設定し、今回の処理で算出された平均Ａｃｃｕｒａｃｙ（Ｍａ）の値を、１ループ前の分割数の時に算出された平均Ａｃｃｕｒａｃｙ（Ｍａ´）として設定し、また、今回の処理で見積もられた残り時間（Ｔｒ）の値を、１ループ前の分割数の時に見積もられた残り時間（Ｔｒ´）として設定する（ステップＳ２１２）。その後、処理はステップＳ２０３へ戻り、新たな分割数（ｎ）を用いた検体データの分割が行われる。

[0067] ステップＳ２１３では、新検知モデルが決定される。ステップＳ２０６における判定処理で、所定以上の差がないと判定された場合（ステップＳ２０６のＮＯ）、検知モデル生成部２６は、平均ＡｃｃｕｒａｃｙがＭａ´の検知モデル群のうち、最大のＡｃｃｕｒａｃｙを持つ検知モデルを新検知モデ

ル１として採用し、当該新検知モデルに対応する機械学習用検知データ及びテスト用検知データの分割パターンを採用する。即ち、検知モデル生成部２６は、これ以上分割数（換言すれば、機械学習用検体データの量）を増やすことによる平均Ａｃｃｕｒａｃｙの向上可能性は低く、１ループ前の分割数での検体データ分割が適正であると判断し、１ループ前の分割数での検体データ分割による機械学習用検体データの量を用いて生成された検知モデルを、新検知モデル１として採用する。その後、本フローチャートに示された処理は終了する（分割成功）。

[0068] なお、上記説明したフローチャートでは、ステップＳ２０９またはステップＳ２１１で条件「Ｎｍａｘ ＜ ｎ」や「Ｔｒ ＜ Ｔｒ´ ／ ２」を満たす場合、「分割失敗」として処理を終了しているが、上述の通り、検体データ分割部２５および検知モデル生成部２６は、処理見積時間が残り時間を超える場合、ここまでの処理で生成された検知モデルを用いて新検知モデルを決定してもよい。これは、これまで実行した結果の範囲でも、現検知モデルと同等以上の新検知モデルを生成できる可能性があるためである。この場合、処理はステップＳ２１３へ進み、検知モデル生成部２６は、ここまでの処理で最大のＡｃｃｕｒａｃｙを持つ検知モデルを新検知モデルとして採用し、対応する分割パターンを採用する。

[0069] 図７は、本実施形態に係る検体データ分割処理（サブ処理Ａ）の流れの概要を示すフローチャートである。本フローチャートは、検体データ分割処理（メイン処理）（図６を参照）のステップＳ２０３の処理を詳細に説明するためのものである。

[0070] ステップＳ３０１では、新たな検体データの数が判定される。新たな検体データの数が分割数（ｎ）よりも少ない場合、分割は実行できないため、本フローチャートに示された処理は終了する（分割失敗）。一方、新たな検体データの数が分割数（ｎ）以上である場合、処理はステップＳ３０２へ進む。

[0071] ステップＳ３０２では、１分割グループあたりの検体データ数が算出され

る。検体データ分割部２５は、「新たな検体データ」の数を分割数（ｎ）で割ることで、１分割グループあたりの「新たな検体データ」の数を算出する。その後、処理はステップＳ３０３へ進む。

[0072] ステップＳ３０３では、各分割グループに、検体データが振り分けられる。検体データ分割部２５は、分割グループ毎に、ステップＳ３０２で算出された１分割グループあたりの「新たな検体データ」の数だけ、検体データをランダムに選択することで、ｎ個の分割グループを生成する。この際、各分割グループには、検体データが重複しないように割り当てられるものとする。その後、本フローチャートに示された処理は終了する（分割成功）。

[0073] 図８は、本実施形態に係る検体データ分割処理（サブ処理Ｂ）の流れの概要を示すフローチャートである。本フローチャートは、検体データ分割処理（メイン処理）（図６を参照）のステップＳ２０５の処理を詳細に説明するためのものである。

[0074] ステップＳ４０１では、変数が初期化される。検体データ分割部２５は、処理に用いる各変数に初期値を設定する。具体的には、検体データ分割部２５は、インデックス（ｉ）に１を設定する。その後、処理はステップＳ４０２へ進む。

[0075] ステップＳ４０２では、分割グループがテスト用と機械学習用とに振り分けられる。検体データ分割部２５は、ｎ個に分割済みの「新たな検体データ」のうち、ｉ番目の分割グループをテスト用に、残りを機械学習用にそれぞれ選択する。その後、処理はステップＳ４０３へ進む。

[0076] ステップＳ４０３では、検知モデルが生成される。検知モデル生成部２６は、変換済みの機械学習用の全検体データと対応する正解ラベルとを入力データとして、機械学習アルゴリズム（例えば、ランダムフォレスト）を使用して検知モデルを生成する。その後、処理はステップＳ４０４へ進む。

[0077] ステップＳ４０４では、Ａｃｃｕｒａｃｙが算出される。検知モデル評価部２８は、変換済みのテスト用の「新たな検体データ」を検知モデルに入力し、Ａｃｃｕｒａｃｙを求める。本実施形態では、Ａｃｃｕｒａｃｙは、以

下の数式を用いて算出される。

Ａｃｃｕｒａｃｙ　＝　（ＴＰ　＋　ＴＮ）　／　（ＴＰ　＋　ＦＮ　＋　ＦＰ　＋　ＴＮ）

ＴＰ：　サイバー攻撃「有」のテスト用検体データのうち、テスト対象の検知モデルを用いて検知された検知データの数

ＦＮ：　サイバー攻撃「有」のテスト用検体データのうち、テスト対象の検知モデルを用いて検知されなかった検知データの数

ＦＰ：　サイバー攻撃「無」のテスト用検体データのうち、テスト対象の検知モデルを用いてサイバー攻撃であると誤検知された検知データの数

ＴＮ：　サイバー攻撃「無」のテスト用検体データのうち、テスト対象の検知モデルを用いて検知されなかった検知データの数

但し、Ａｃｃｕｒａｃｙは、その他の方法で算出されてもよい。その後、処理はステップＳ４０５へ進む。

[0078] ステップＳ４０５およびステップＳ４０６では、インデックス（ｉ）が更新され、処理を継続するか否かの判定が行われる。検体データ分割部２５は、インデックス（ｉ）を１インクリメントする（ステップＳ４０５）。そして、更新されたインデックス（ｉ）は、分割数（ｎ）と比較され、比較の結果インデックス（ｉ）が分割数（ｎ）以下である場合（ステップＳ４０６のＮＯ）、処理はステップＳ４０２へ戻り、ステップＳ４０２からステップＳ４０５の処理が再実行される。即ち、ステップＳ４０２からステップＳ４０５の処理は、分割数（ｎ）だけ繰り返し実行され、ｎ個の検知モデルが生成される。一方、比較の結果インデックス（ｉ）が分割数（ｎ）より大きい場合（ステップＳ４０６のＹＥＳ）、処理はステップＳ４０７へ進む。

[0079] ステップＳ４０７では、平均Ａｃｃｕｒａｃｙが算出される。検体データ分割部２５は、生成されたｎ個の検知モデルのＡｃｃｕｒａｃｙから、平均Ａｃｃｕｒａｃｙを求める。その後、本フローチャートに示された処理は終了する。

[0080] 図９は、本実施形態に係る第二の新検知モデル生成処理の流れの概要を示

すフローチャートである。本フローチャートは、更新用検知モデル生成処理（図４を参照）のステップＳ００６の処理を詳細に説明するためのものである。

[0081] ステップＳ５０１では、既存の特徴および新規の特徴を使用した検体データの変換が行われる。検体データ変換部２４は、既存の特徴および新規の特徴を使用して、全検体データ（既存の検体データ＋新たな検体データ）を、機械学習アルゴリズムの入力データに適した形式に変換する。その後、処理はステップＳ５０２へ進む。

[0082] ステップＳ５０２では、第二の新検知モデルが生成される。第二の検知モデル生成部２７は、変換したデータおよび対応する正解ラベル（サイバー攻撃の有無を示す情報）を入力データとして、機械学習アルゴリズムを使用して検知モデルを生成する。この際用いられる機械学習アルゴリズムおよび機械学習用パラメータの少なくとも一方は、ステップＳ４０３における検知モデル生成で用いられたものと異なる。その後、本フローチャートに示された処理は終了する。

[0083] 図１０は、本実施形態に係る性能比較処理（メイン処理）の流れの概要を示すフローチャートである。本フローチャートは、更新用検知モデル生成処理（図４を参照）のステップＳ００７の処理を詳細に説明するためのものである。なお、本実施形態に係るステップＳ６０１およびステップＳ６０２の処理は、単一のサーバーまたは複数のサーバーで並列に処理される。

[0084] ステップＳ６０１では、新たな検体データをテスト用検体データとして用いて、現検知モデルとの性能（検知精度）比較が行われる。性能比較部２９は、変換済みの新たな検体データ（テスト用）を使って、現検知モデルの性能と同等以上の新検知モデル（第二の新検知モデルを含む）が存在するか確認する。性能比較処理（サブ処理）の詳細については、図１１を用いて後述する。なお、ここで抽出された新検知モデルを、「候補Ａ」とする。その後、処理はステップＳ６０３へ進む。

[0085] ステップＳ６０２では、検体データ管理データベースに蓄積されていた既

存のテスト用検体データのうち、環境識別子が「共通」である検体データおよび「環境Ａ」である検体データを用いて、現検知モデルとの性能比較が行われる。性能比較部２９は、変換済みの既存の検体データ（テスト用）を使って、現検知モデルの性能と同等以上の新検知モデル（第二の新検知モデルを含む）が存在するか確認する。性能比較処理（サブ処理）の詳細については、図１１を用いて後述する。なお、ここで抽出された新検知モデルを、「候補Ｂ」とする。その後、処理はステップＳ６０３へ進む。

[0086] ステップＳ６０３では、「候補Ａ」および「候補Ｂ」が決定されたか否かが判定される。性能比較処理（サブ処理）の結果、「候補Ａ」に該当する新検知モデルおよび「候補Ｂ」に該当する新検知モデルが何れか一方でも存在しない場合、本フローチャートに示された処理は終了する（置き換え対象なし）。一方、性能比較処理（サブ処理）の結果、「候補Ａ」に該当する新検知モデルおよび「候補Ｂ」に該当する新検知モデルが存在する場合、処理はステップＳ６０４へ進む。

[0087] ステップＳ６０４からステップＳ６０６では、更新用検知モデルが決定される。「候補Ａ」と「候補Ｂ」とが同一の新検知モデルである場合、性能比較部２９は、当該新検知モデルを、更新用検知モデルとして採用する（ステップＳ６０５）。一方、「候補Ａ」と「候補Ｂ」とが異なる新検知モデルである場合、性能比較部２９は、「候補Ａ」の新検知モデルを、更新用検知モデルとして採用する（ステップＳ６０６）。これは、新たな検体データに係る攻撃等を検知できる検知モデルを優先的に採用するためである。その後、本フローチャートに示された処理は終了する（置き換え対象決定）。

[0088] 図１１は、本実施形態に係る性能比較処理（サブ処理）の流れの概要を示すフローチャートである。本フローチャートは、性能比較処理（メイン処理）（図１０を参照）のステップＳ６０１またはステップＳ６０２の処理を詳細に説明するためのものである。

[0089] ステップＳ７０１では、性能比較用の評価値が算出される。性能比較部２９は、変換済みのテスト用検体データを使用して、現検知モデルおよび新検

知モデルのサイバー攻撃の検知率（以下、「Ｒｅｃａｌｌ」と称する）および検知結果の正答率（以下、「Ｐｒｅｃｉｓｉｏｎ」と称する）を求める。本実施形態では、ＲｅｃａｌｌおよびＰｒｅｃｉｓｉｏｎは、以下の数式を用いて算出される。

Ｒｅｃａｌｌ　＝　ＴＰ　／　（ＴＰ　＋　ＦＮ）

Ｐｒｅｃｉｓｉｏｎ　＝　ＴＰ　／　（ＴＰ　＋　ＦＰ）

ＴＰ：　サイバー攻撃「有」のテスト用検体データのうち、テスト対象の検知モデルを用いて検知された検知データの数

ＦＮ：　サイバー攻撃「有」のテスト用検体データのうち、テスト対象の検知モデルを用いて検知されなかった検知データの数

ＦＰ：　サイバー攻撃「無」のテスト用検体データのうち、テスト対象の検知モデルを用いてサイバー攻撃であると誤検知された検知データの数

ＴＮ：　サイバー攻撃「無」のテスト用検体データのうち、テスト対象の検知モデルを用いて検知されなかった検知データの数

但し、ＲｅｃａｌｌおよびＰｒｅｃｉｓｉｏｎは、その他の方法で算出されてもよい。その後、処理はステップＳ７０２へ進む。

[0090] ステップＳ７０２では、算出された評価値を用いた性能比較が行われる。具体的には、性能比較部２９は、算出された評価値が「現検知モデルのＲｅｃａｌｌ　＜＝　新検知モデルのＲｅｃａｌｌかつ、現検知モデルのＰｒｅｃｉｓｉｏｎ　＜＝　新検知モデルのＰｒｅｃｉｓｉｏｎ」という条件を満たすか否かを判定する。判定の結果、評価値が条件を満たさない場合、本フローチャートに示された処理は終了する（置き換え候補なし）。一方、判定の結果、評価値が条件を満たす場合、処理はステップＳ７０３へ進む。

[0091] ステップＳ７０３からステップＳ７０５では、条件を満たす新検知モデルから、置き換え候補が決定される。ここで決定される候補は、本フローチャートの処理がステップＳ６０１から呼び出されている場合には「候補Ａ」、本フローチャートの処理がステップＳ６０２から呼び出されている場合には「候補Ｂ」である。条件を満たす新検知モデルが複数存在する場合（ステッ

プＳ７０３のＹＥＳ）、性能比較部２９は、Ｐｒｅｃｉｓｉｏｎが最も大きい新検知モデルを、最も高い性能を持つ検知モデルとして採用し、置き換え候補（「候補Ａ」または「候補Ｂ」）とする（ステップＳ７０４）。一方、条件を満たす新検知モデルが１つである場合、性能比較部２９は、該当する新検知モデルを、最も高い性能を持つ検知モデルとして採用し、置き換え候補（「候補Ａ」または「候補Ｂ」）とする。その後、本フローチャートに示された処理は終了する（置き換え候補決定）。

[0092] 即ち、本実施形態において、性能比較部２９は、所定の種類の通信の検知率（実際の攻撃が攻撃として検知される割合）および所定の種類の通信であるとの検知結果の正答率（攻撃として検知されたものが、実際の攻撃である割合）の双方に基づいて、性能比較を行う。これは、正常な通信をサイバー攻撃と検知してしまった場合、正常な通信が制限、または、遮断されることにつながり、運用継続を維持できなくなるため、サイバー攻撃を検知できる能力が高いことが重要であると同時に、正常な通信をサイバー攻撃と誤認することも避ける必要があるためである。本実施形態によれば、上記のような性能比較が行われるため、安全性の向上の側面からの検知漏れの評価（検知率）と、運用継続性の側面からの正確性の評価（正答率）との双方の側面からの性能比較が行われ、現検知モデルの置き換え対象として最適な新検知モデルを選択することが可能となる。

[0093] ＜効果＞

本実施形態に係る情報処理装置、方法およびプログラムによれば、検知システムの検知ロジック更新を省力化できると共に、継続的な更新が容易になる。即ち、限られた検体データを用いて、性能の高い検知モデルを効率的に生成することが可能となり、検知ロジックを更新する手間を軽減することができる。また、「運用継続性の維持」と「安全性の向上」とを両立した検知モデルを生成可能であることから、サービスレベルの維持に加えて、サイバー攻撃に対する安全性の向上にも対応できる。

符号の説明

[0094] １　システム

１０　検知モデル生成装置

２０　ネットワーク監視装置

９０　ノード

請求の範囲

[請求項1] 所定の種類の通信に係る新たな検体データの入力を受け付ける検体データ入力受付手段と、

前記新たな検体データを、互いに異なる複数の分割パターンを用いて機械学習用検体データとテスト用検体データに分割する検体データ分割手段と、

前記機械学習用検体データを用いた機械学習によって、分割パターン毎に、前記所定の種類の通信を検知するための検知モデルを生成する検知モデル生成手段と、

前記テスト用検体データを用いて前記検知モデルの性能を評価する検知モデル評価手段と、を備え、

前記検知モデル生成手段は、生成された検知モデルのうち最も性能の高いモデルを新検知モデルとする、

情報処理装置。

[請求項2] 前記検体データ分割手段は、第一の分割パターンを用いて生成された検知モデルの性能と、該第一の分割パターンより機械学習用検体データを増やした第二の分割パターンを用いて生成された検知モデルの性能との間に所定以上の差がある場合、該第二の分割パターンよりも機械学習用検体データを増やした第三の分割パターンを更に生成する、

請求項１に記載の情報処理装置。

[請求項3] 前記検体データ分割手段は、第一の分割パターンを用いて生成された検知モデルの性能に比べて、該第一の分割パターンより機械学習用検体データを増やした第二の分割パターンを用いて生成された検知モデルの性能が所定以上優れている場合、該第二の分割パターンよりも機械学習用検体データを増やした第三の分割パターンを更に生成する、

請求項１に記載の情報処理装置。

[請求項4] 前記所定の種類の通信を検知するための現検知モデルの生成に用いられた既存の特徴および既存の検体データを保存するデータ管理手段を更に備え、

前記検知モデル生成手段は、前記機械学習用検体データおよび前記既存の検体データを用いた機械学習によって分割パターン毎に検知モデルを生成する、

請求項１から３の何れか一項に記載の情報処理装置。

[請求項5] 前記新たな検体データから、前記既存の特徴とは異なる新規の特徴を抽出する特徴抽出手段と、

前記既存の特徴および前記新規の特徴を使用して、既存の検体データおよび前記新たな検体データを、機械学習のための所定のデータ形式に変換する、検体データ変換手段を更に備え、

前記検知モデル生成手段は、前記所定のデータ形式に変換された検体データを用いた機械学習によって検知モデルを生成する、

請求項４に記載の情報処理装置。

[請求項6] 前記データ管理手段は、前記新たな検体データおよび前記新規の特徴を、前記新検知モデルの生成後に、既存の検体データおよび既存の特徴として蓄積する、

請求項５に記載の情報処理装置。

[請求項7] 前記検体データ入力受付手段は、前記所定の種類の通信を検知するための現検知モデルが用いられている所定の環境に係る新たな検体データの入力を受け付け、

前記データ管理手段は、前記新検知モデルの生成後に、前記新たな検体データおよび前記新規の特徴を、前記所定の環境に関連づけて蓄積することで、該所定の環境に関する既存の検体データおよび既存の特徴とする、

請求項５に記載の情報処理装置。

[請求項8] 検体データ変換手段は、Ｂａｇ－ｏｆ－Ｗｏｒｄｓ方式で検体デー

タを変換し、

検知モデル生成手段は、ランダムフォレスト方式の機械学習によって検知モデルを生成する、

請求項５から７の何れか一項に記載の情報処理装置。

[請求項9] 検体データ変換手段は、検体データ内のデータを各特徴に対応する一意の数値に変換する方法で検体データを変換し、

検知モデル生成手段は、ニューラルネットワーク方式の機械学習によって検知モデルを生成する、

請求項５から７の何れか一項に記載の情報処理装置。

[請求項10] 前記所定の種類の通信を検知するための現検知モデルと前記新検知モデルの性能を比較し、現検知モデルと概略同等かまたは現検知モデルよりも大きい性能を有する前記新検知モデルを、現検知モデルを置き換える更新用検知モデルに決定する性能比較手段を更に備える、

請求項１から９の何れか一項に記載の情報処理装置。

[請求項11] 前記検知モデル生成手段とは異なる機械学習アルゴリズムまたは異なる機械学習用パラメータを用いて第二の新検知モデルを生成する第二の検知モデル生成手段を更に備え、

前記性能比較手段は、前記現検知モデル、前記新検知モデルおよび前記第二の新検知モデルの性能を比較し、現検知モデルと概略同等かまたは現検知モデルよりも大きい性能を有する何れかの前記新検知モデルまたは前記第二の新検知モデルを、現検知モデルを置き換える更新用検知モデルとする、

請求項１０に記載の情報処理装置。

[請求項12] 前記性能比較手段は、前記所定の種類の通信の検知率および前記所定の種類の通信であるとの検知結果の正答率の双方に基づいて、前記性能比較を行う、

請求項１０または１１に記載の情報処理装置。

[請求項13] 前記更新用検知モデルを検知装置に送信し、該検知装置において用

いられる検知モデルを更新させる、検知モデル送信手段を更に備える、

請求項１０から１２の何れか一項に記載の情報処理装置。

[請求項14] 前記検体データ分割手段が異なる分割パターンを用いて前記新たな検体データを分割する毎に、前記更新用検知モデルが得られるまでの予め設定された上限時間までの残り時間と処理見積時間を計算し、該処理見積時間が残り時間を超えるか否かを判定する残り時間判定手段を更に備え、

前記検体データ分割手段は、前記処理見積時間が残り時間を超える場合、処理を終了する、

請求項１０から１３の何れか一項に記載の情報処理装置。

[請求項15] 前記検知モデル生成手段は、前記処理見積時間が残り時間を超える場合、これまでの処理で生成された検知モデルを用いて前記新検知モデルを決定する、

請求項１４に記載の情報処理装置。

[請求項16] 前記所定の種類の通信は、所定の種類の悪意の通信である、

請求項１から１４の何れか一項に記載の情報処理装置。

[請求項17] コンピューターが、

所定の種類の通信に係る新たな検体データの入力を受け付ける検体データ入力受付と、

前記新たな検体データを、互いに異なる複数の分割パターンを用いて機械学習用検体データとテスト用検体データに分割する検体データ分割と、

前記機械学習用検体データを用いた機械学習によって、分割パターン毎に、前記所定の種類の通信を検知するための検知モデルを生成する検知モデル生成と、

前記テスト用検体データを用いて前記検知モデルの性能を評価する検知モデル評価と、を実行し、

前記検知モデル生成では、生成された検知モデルのうち最も性能の高いモデルが新検知モデルとされる、

方法。

[請求項18] コンピューターを、

所定の種類の通信に係る新たな検体データの入力を受け付ける検体データ入力受付手段と、

前記新たな検体データを、互いに異なる複数の分割パターンを用いて機械学習用検体データとテスト用検体データに分割する検体データ分割手段と、

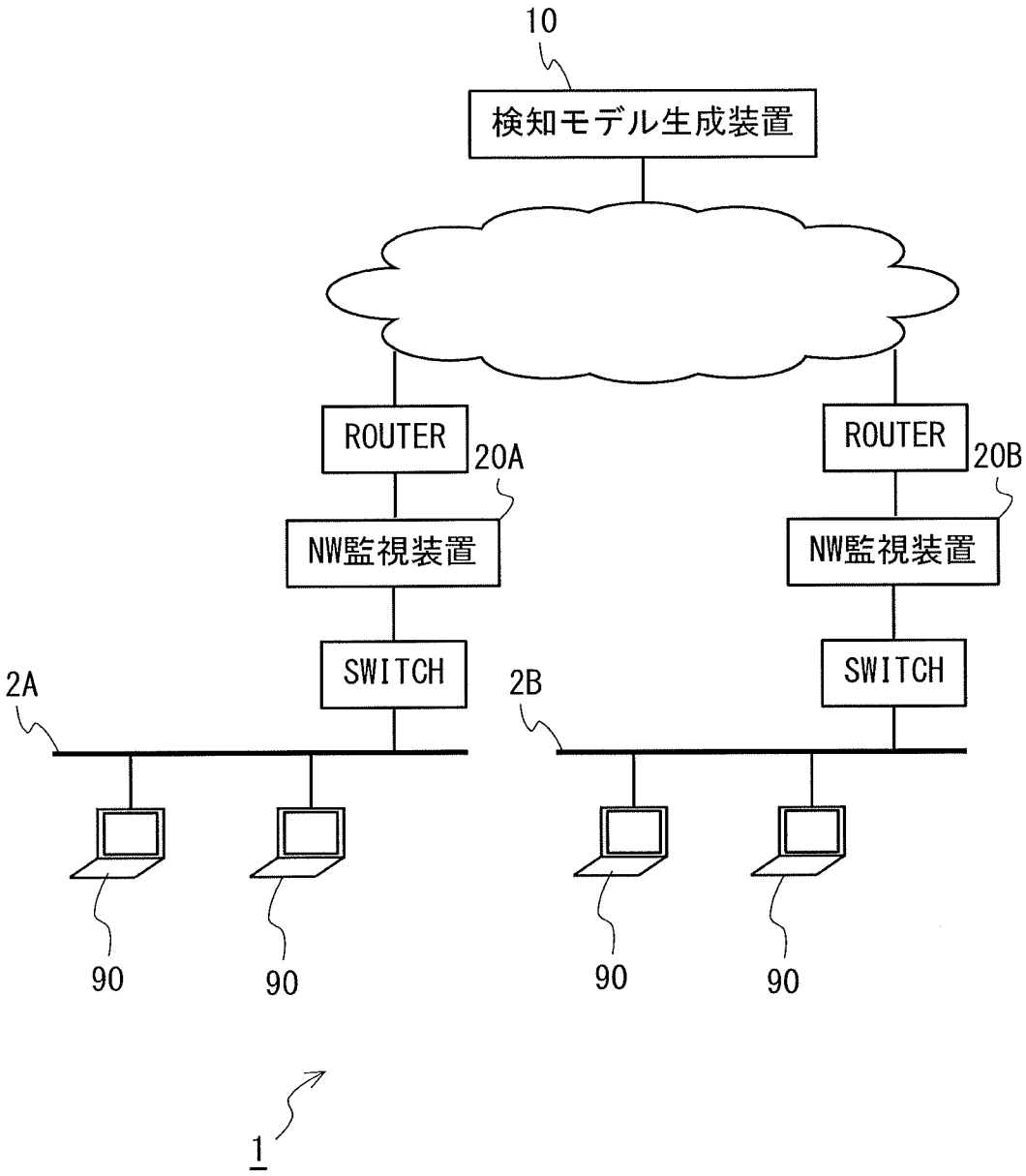
前記機械学習用検体データを用いた機械学習によって、分割パターン毎に、前記所定の種類の通信を検知するための検知モデルを生成する検知モデル生成手段と、

前記テスト用検体データを用いて前記検知モデルの性能を評価する検知モデル評価手段と、として機能させ、

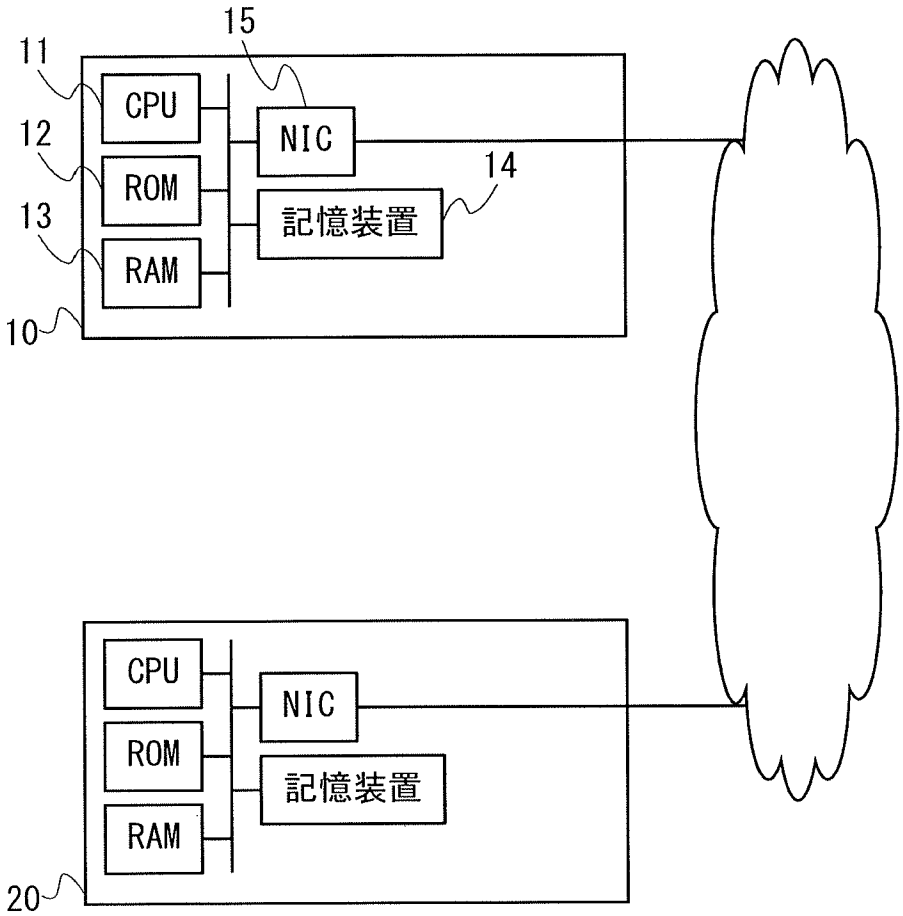
前記検知モデル生成手段は、生成された検知モデルのうち最も性能の高いモデルを新検知モデルとする、

プログラム。

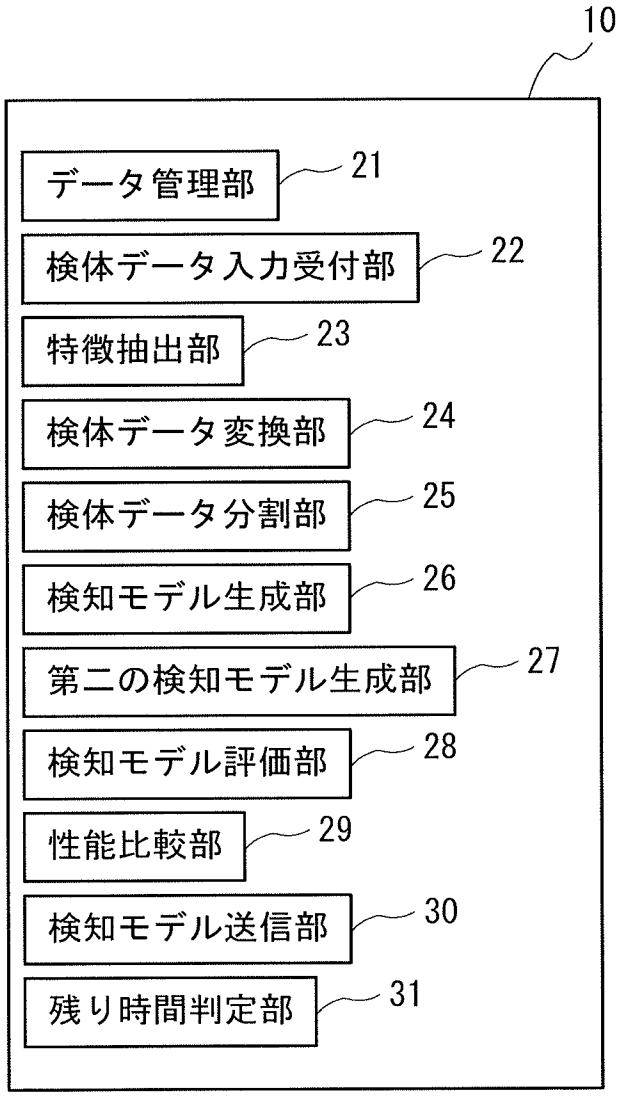
[図1]


10
検知モデル生成装置
ROUTER
20A
ROUTER
20B
NW監視装置
NW監視装置
SWITCH
SWITCH
2A
2B
90
90
90
90
1

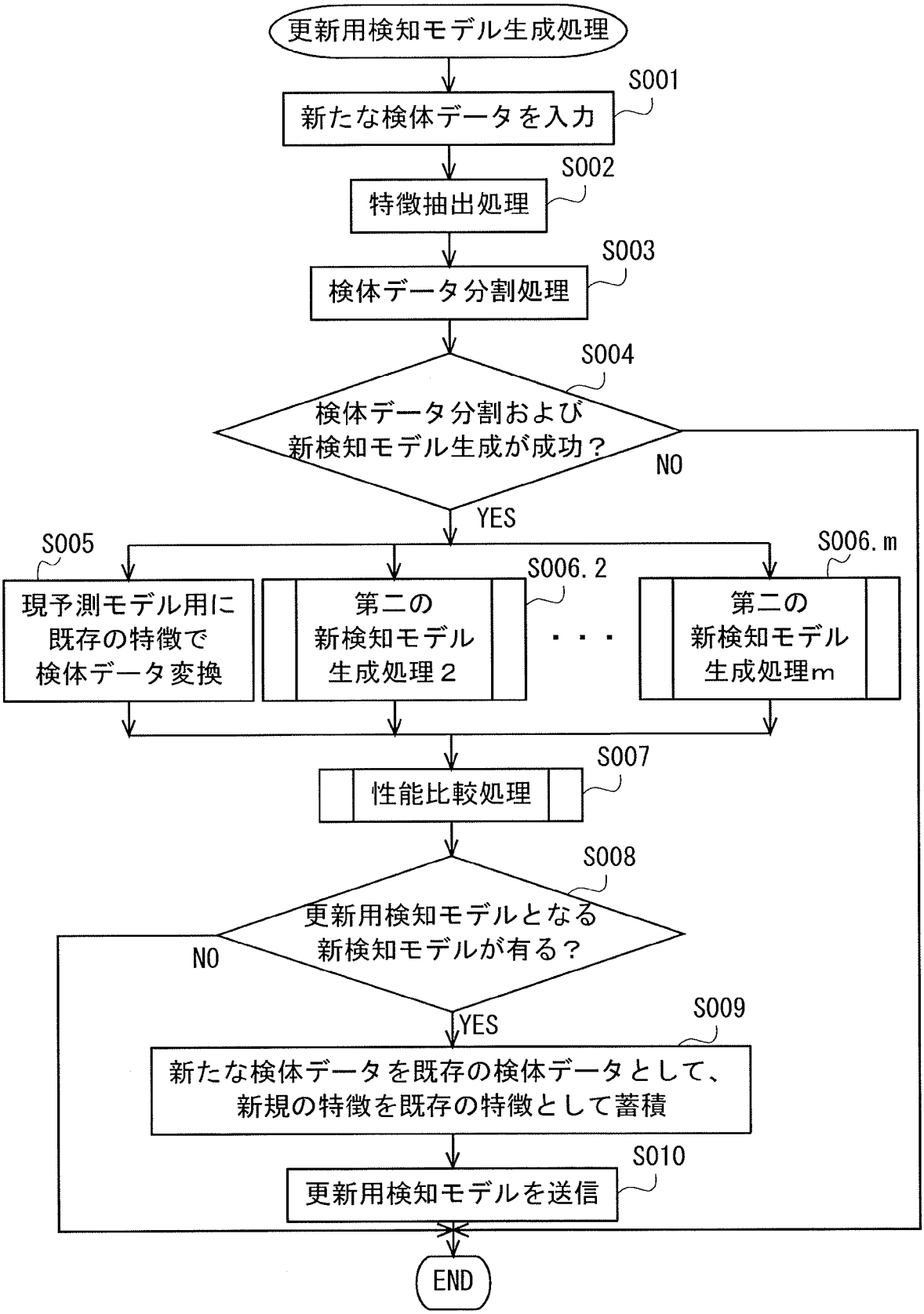
[図2]



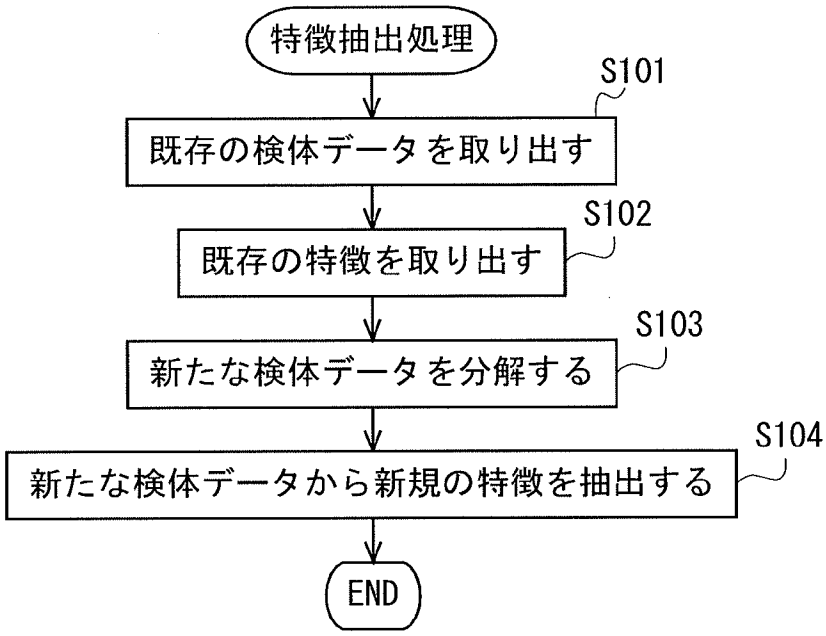
[図3]


10
データ管理部
21
検体データ入力受付部
22
特徴抽出部
23
検体データ変換部
24
検体データ分割部
25
検知モデル生成部
26
第二の検知モデル生成部
27
検知モデル評価部
28
性能比較部
29
検知モデル送信部
30
残り時間判定部
31

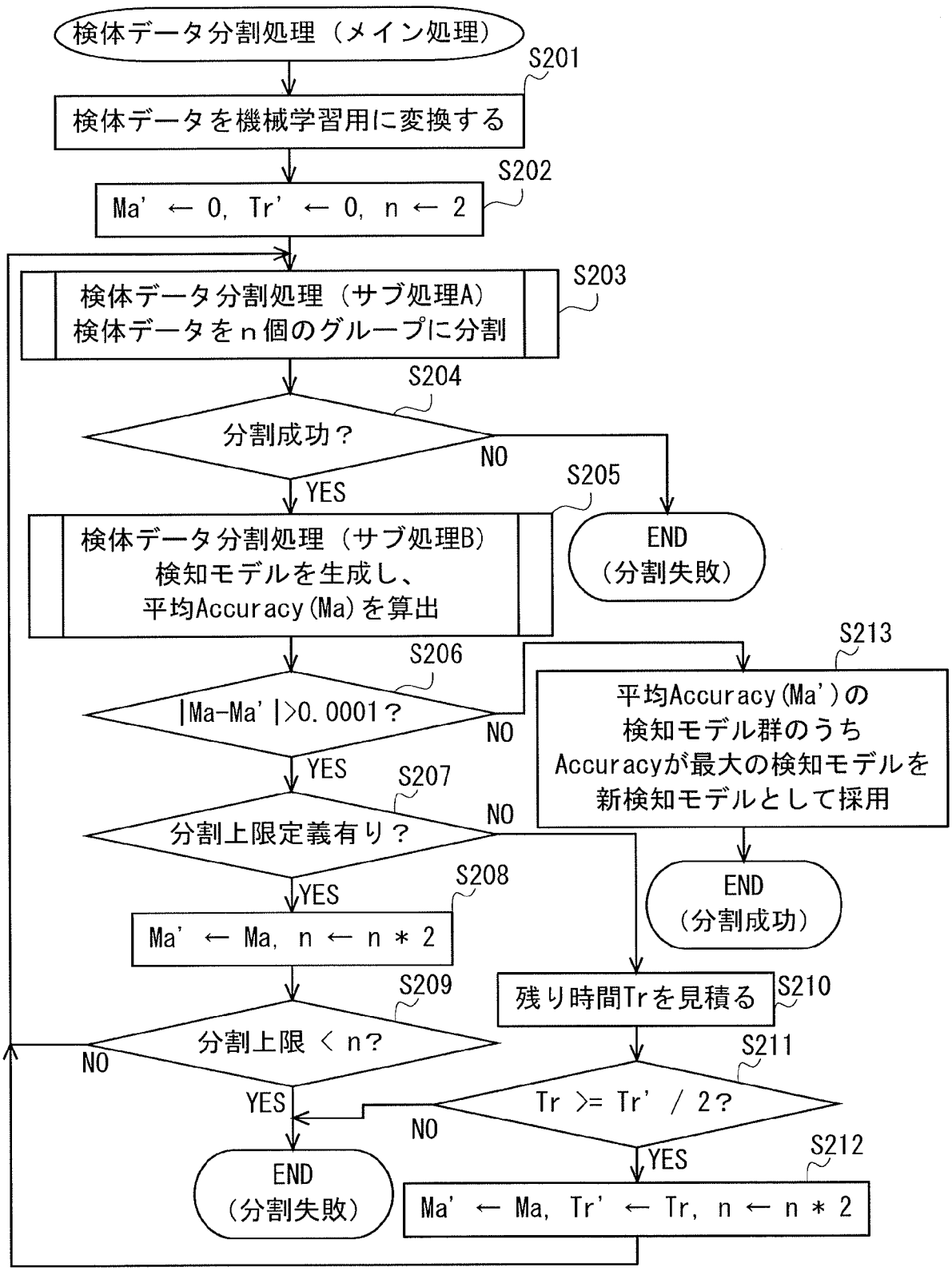
[図4]


更新用検知モデル生成処理
S001
新たな検体データを入力
S002
特徴抽出処理
S003
検体データ分割処理
S004
検体データ分割および
新検知モデル生成が成功？
NO
YES
S005
現予測モデル用に
既存の特徴で
検体データ変換
S006.2
第二の
新検知モデル
生成処理２
・・・
S006.m
第二の
新検知モデル
生成処理m
S007
性能比較処理
S008
更新用検知モデルとなる
新検知モデルが有る？
NO
YES
S009
新たな検体データを既存の検体データとして、
新規の特徴を既存の特徴として蓄積
S010
更新用検知モデルを送信
END

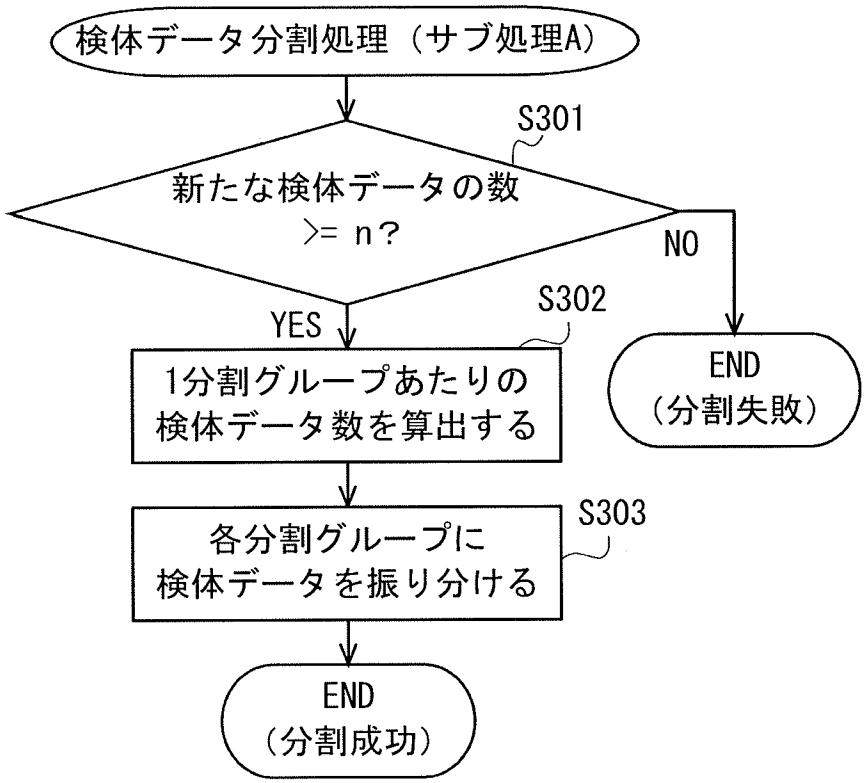
[図5]


特徴抽出処理
S101
既存の検体データを取り出す
S102
既存の特徴を取り出す
S103
新たな検体データを分解する
S104
新たな検体データから新規の特徴を抽出する
END

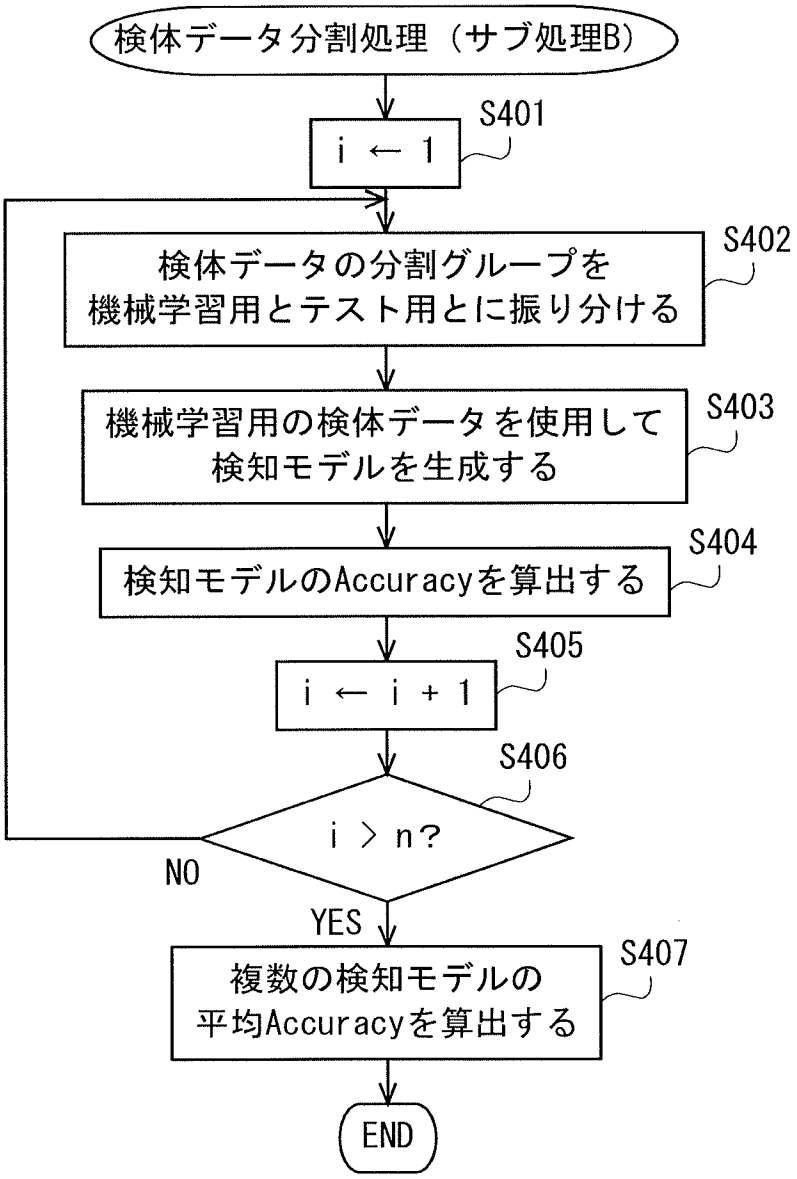
[図6]


検体データ分割処理（メイン処理）
S201
検体データを機械学習用に変換する
S202
Ma' ← 0, Tr' ← 0, n ← 2
S203
検体データ分割処理（サブ処理A）
検体データをｎ個のグループに分割
S204
分割成功？
NO
YES
S205
検体データ分割処理（サブ処理B）
検知モデルを生成し、
平均Accuracy(Ma)を算出
END
（分割失敗）
S213
S206
|Ma-Ma'|>0.0001？
NO
YES
平均Accuracy(Ma')の
検知モデル群のうち
Accuracyが最大の検知モデルを
新検知モデルとして採用
S207
分割上限定義有り？
NO
YES
END
（分割成功）
S208
Ma' ← Ma, n ← n * 2
S209
分割上限 < n?
NO
YES
残り時間Trを見積る
S210
S211
Tr >= Tr' / 2?
NO
YES
END
（分割失敗）
S212
Ma' ← Ma, Tr' ← Tr, n ← n * 2

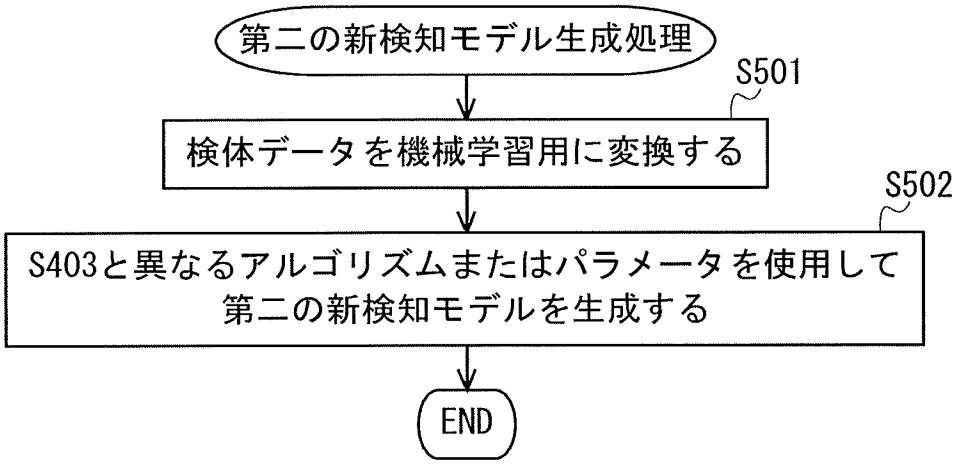
[図7]


検体データ分割処理（サブ処理A）
S301
新たな検体データの数
>= n？
NO
YES
S302
END
（分割失敗）
1分割グループあたりの
検体データ数を算出する
S303
各分割グループに
検体データを振り分ける
END
（分割成功）

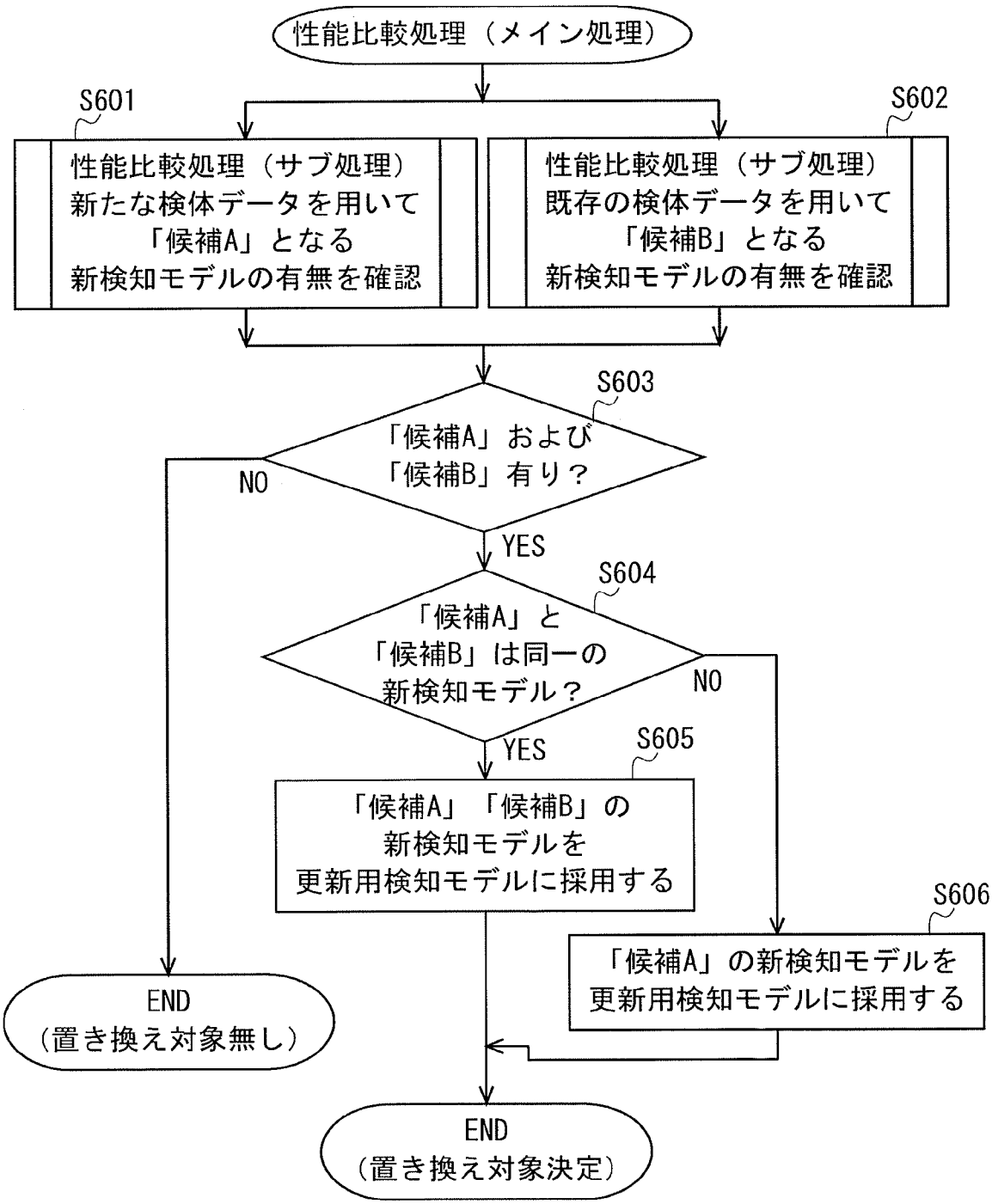
[図8]


検体データ分割処理（サブ処理B）
S401
i ← 1
S402
検体データの分割グループを
機械学習用とテスト用とに振り分ける
S403
機械学習用の検体データを使用して
検知モデルを生成する
S404
検知モデルのAccuracyを算出する
S405
i ← i + 1
S406
i > n?
NO
YES
S407
複数の検知モデルの
平均Accuracyを算出する
END

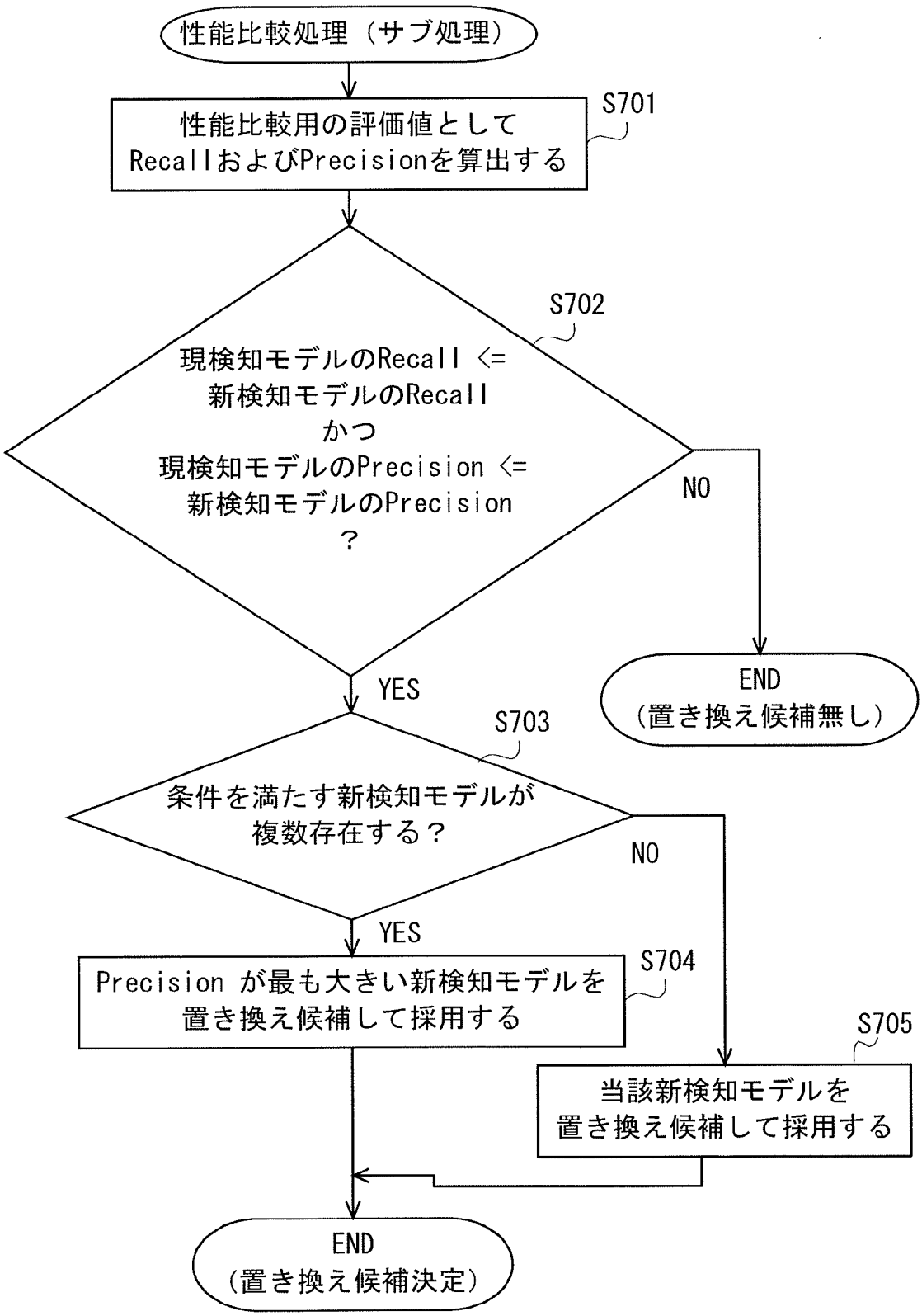
[図9]


第二の新検知モデル生成処理
S501
検体データを機械学習用に変換する
S502
S403と異なるアルゴリズムまたはパラメータを使用して
第二の新検知モデルを生成する
END

[図10]


性能比較処理（メイン処理）
S601
性能比較処理（サブ処理）
新たな検体データを用いて
「候補A」となる
新検知モデルの有無を確認
S602
性能比較処理（サブ処理）
既存の検体データを用いて
「候補B」となる
新検知モデルの有無を確認
S603
「候補A」および
「候補B」有り？
NO
YES
S604
「候補A」と
「候補B」は同一の
新検知モデル？
NO
YES
S605
「候補A」「候補B」の
新検知モデルを
更新用検知モデルに採用する
S606
「候補A」の新検知モデルを
更新用検知モデルに採用する
END
（置き換え対象無し）
END
（置き換え対象決定）

[図11]


性能比較処理（サブ処理）
性能比較用の評価値として
RecallおよびPrecisionを算出する
S701
S702
現検知モデルのRecall <=
新検知モデルのRecall
かつ
現検知モデルのPrecision <=
新検知モデルのPrecision
？
NO
YES
END
（置き換え候補無し）
S703
条件を満たす新検知モデルが
複数存在する？
NO
YES
Precision が最も大きい新検知モデルを
置き換え候補して採用する
S704
S705
当該新検知モデルを
置き換え候補して採用する
END
（置き換え候補決定）

INTERNATIONAL SEARCH REPORT

International application No.
PCT/JP2016/085227

A. CLASSIFICATION OF SUBJECT MATTER
H04L12/66(2006.01)i

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)
H04L12/66

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Jitsuyo Shinan Koho	1922-1996	Jitsuyo Shinan Toroku Koho	1996-2017
Kokai Jitsuyo Shinan Koho	1971-2017	Toroku Jitsuyo Shinan Koho	1994-2017

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 A	US 2015/0135318 A1 (CHI TIN, H et al.), 14 May 2015 (14.05.2015), paragraphs [0015] to [0017]; fig. 1, 2 (Family: none)	1,4-18 2,3
A	US 2010/0082513 A1 (LIU, L), 01 April 2010 (01.04.2010), (Family: none)	1-18
A	US 2015/0193697 A1 (VASSEUR, JP et al.), 09 July 2015 (09.07.2015), & WO 2015/103514 A1 & WO 2015/103520 A1	1-18

☐ Further documents are listed in the continuation of Box C. ☐ 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	Date of mailing of the international search report
07 February 2017 (07.02.17)	14 February 2017 (14.02.17)

Name and mailing address of the ISA/	Authorized officer
Japan Patent Office 3-4-3,Kasumigaseki,Chiyoda-ku, Tokyo 100-8915,Japan	Telephone No.

国際調査報告

国際出願番号　ＰＣＴ／ＪＰ２０１６／０８５２２７

Ａ．　発明の属する分野の分類（国際特許分類（ＩＰＣ））

Int.Cl.　H04L12/66(2006.01)i

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））

Int.Cl.　H04L12/66

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	１９７１－２０１７年
日本国実用新案登録公報	１９９６－２０１７年
日本国登録実用新案公報	１９９４－２０１７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 A	US 2015/0135318 A1 (CHI TIN, H et al.) 2015.05.14, [0015]-[0017], figures 1, 2 (family: none)	1, 4-18 2,3
A	US 2010/0082513 A1 (LIU, L) 2010.04.01, (family: none)	1-18
A	US 2015/0193697 A1 (VASSEUR, JP et al) 2015.07.09, & WO 2015/103514 A1 & WO 2015/103520 A1	1-18

☐　Ｃ欄の続きにも文献が列挙されている。

☐　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矛盾するものではなく、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 ０７．０２．２０１７	国際調査報告の発送日 １４．０２．２０１７
国際調査機関の名称及びあて先 日本国特許庁（ＩＳＡ／ＪＰ） 郵便番号１００－８９１５ 東京都千代田区霞が関三丁目４番３号	特許庁審査官（権限のある職員）　５Ｘ　３３６５ 速水　雄太 電話番号　０３－３５８１－１１０１　内線　３５９６