

(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
2024年3月14日(14.03.2024)



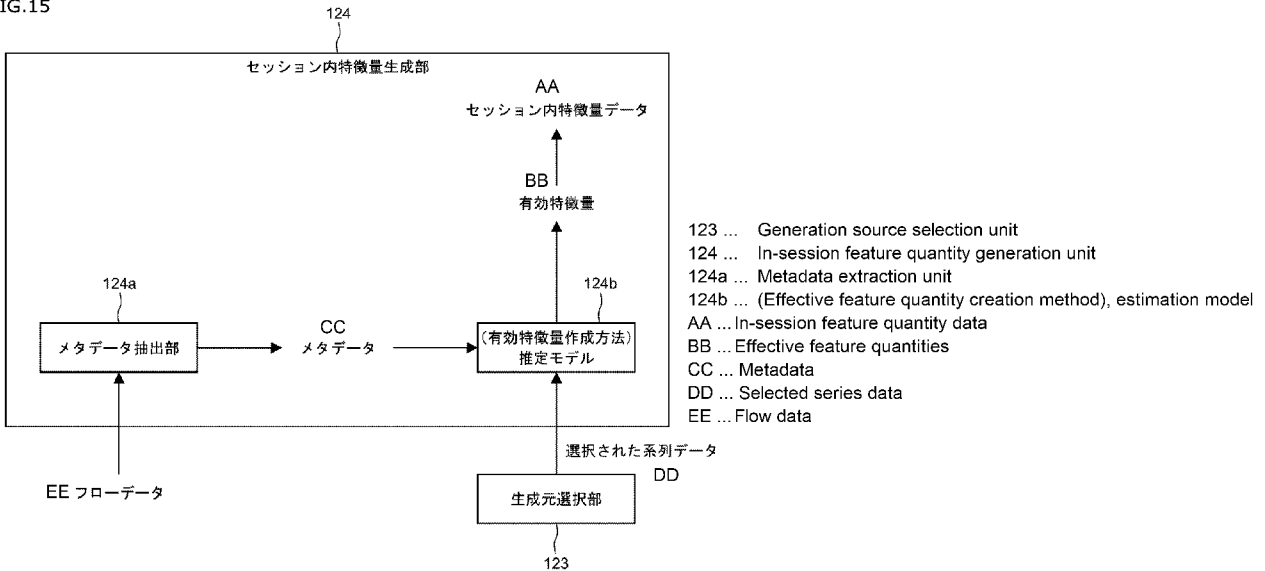
(10) 国際公開番号
WO 2024/053370 A1

- (51) 国際特許分類:
G06N 20/00 (2019.01)
- (21) 国際出願番号: PCT/JP2023/029935
- (22) 国際出願日: 2023年8月21日(21.08.2023)
- (25) 国際出願の言語: 日本語
- (26) 国際公開の言語: 日本語
- (30) 優先権データ:
特願 2022-141171 2022年9月6日(06.09.2022) JP
- (71) 出願人: ソニーグループ株式会社(SONY GROUP CORPORATION) [JP/JP]; 〒1080075 東京都港区港南1丁目7番1号 Tokyo (JP).
- (72) 発明者: 中田 健人(NAKADA Kento); 〒1080075 東京都港区港南1丁目7番1号 ソニーグループ株式会社内 Tokyo (JP). 浅井 智佳子(ASAI Chikako); 〒1080075 東京都港区港南1丁目7番1号 ソニーグループ株式会社内 Tokyo (JP). 高松 慎吾(TAKAMATSU Shingo); 〒1080075 東京都港区港南1丁目7番1号 ソニーグループ株式会社内 Tokyo (JP).
- (74) 代理人: 西川 孝, 外 (NISHIKAWA Takashi et al.); 〒1700013 東京都豊島区東池袋3丁目9番10号 池袋F Nビル4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, CV, CZ, DE, DJ, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IQ, IR, IS, IT, JM, JO, JP, KE, KG, KH, KN, KP, KR, KW, KZ, LA, LC, LK, LR, LS, LU, LY, MA, MD, MG, MK, MN, MU, 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,

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

(54) 発明の名称: 情報処理装置、および情報処理方法、並びにプログラム

FIG. 15



(57) Abstract: The present disclosure relates to an information processing device, an information processing method, and a program which make it possible to efficiently find and extract, from time-series data, feature quantities effective in creating a machine learning model. The present disclosure involves: generating metadata from flow data including at least time-series data; estimating a method for generating feature quantities from series data constituting the flow data, on the basis of the generated metadata; and generating feature quantities from the series data using the estimated generation method. The present disclosure can be applied to technology for generating feature quantities necessary to train machine learning models.

WO 2024/053370 A1

SL, 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, CV, GH, GM, KE, LR, LS, MW, MZ, NA, RW, SC, 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, ME, 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:

- 国際調査報告 (条約第21条(3))
-

(57) 要約: 本開示は、時系列データから機械学習モデルの作成において有効な特徴量を効率的に探索して抽出できるようにする情報処理装置、および情報処理方法、並びにプログラムに関する。少なくとも時系列データを含むフローデータからメタデータを生成し、生成したメタデータに基づいて、フローデータを構成する系列データより特徴量の生成方法を推定し、推定された生成方法で、系列データより特徴量を生成する。機械学習モデルの学習に必要なとされる特徴量を生成する技術に適用することができる。

明 細 書

発明の名称：

情報処理装置、および情報処理方法、並びにプログラム

技術分野

[0001] 本開示は、情報処理装置、および情報処理方法、並びにプログラムに関し、特に、時系列データから機械学習モデルの作成において有効な特徴量を効率的に探索して抽出できるようにした情報処理装置、および情報処理方法、並びにプログラムに関する。

背景技術

[0002] IoT (Internet of Things) においては複数の時系列データからなるデータ集合が蓄積されることが多くなってきている。

[0003] 一方、その様なデータを用いた機械学習モデルや因果モデルの構築は高度な専門性が求められるため、専門性の乏しい人でもモデルを構築可能なツールが期待されている。

[0004] そこで、時系列データから目的事象の発生を予測する機械学習モデルを生成する際に、機械学習の入力データ数が膨大になるのを抑制し、負例の時系列データの基準日を決定するものが提案されている（特許文献1参照）。

先行技術文献

特許文献

[0005] 特許文献1：特開2021-189833号公報

発明の概要

発明が解決しようとする課題

[0006] しかしながら、特許文献1の技術においては、問題設定が限定されていると共に、前処理である特徴量を生成する作業が非常に煩雑であった。

[0007] 本開示は、このような状況に鑑みてなされたものであり、特に、時系列データから機械学習モデルの作成において有効な特徴量を効率的に探索して抽出できるようにするものである。

課題を解決するための手段

- [0008] 本開示の一側面の情報処理装置、およびプログラムは、少なくとも時系列データを含むフローデータのメタデータを生成するメタデータ生成部と、前記メタデータに基づいて、前記フローデータを構成する系列データより特徴量の生成方法を推定する推定部と、前記推定部により推定された生成方法で、前記系列データより特徴量を生成する特徴量生成部とを備える情報処理装置、およびプログラムである。
- [0009] 本開示の一側面の情報処理方法は、少なくとも時系列データを含むフローデータのメタデータを生成し、前記メタデータに基づいて、前記フローデータを構成する系列データより特徴量の生成方法を推定し、推定された生成方法で、前記系列データより特徴量を生成するステップを含む情報処理方法である。
- [0010] 本開示の一側面においては、少なくとも時系列データを含むフローデータのメタデータが生成され、前記メタデータに基づいて、前記フローデータを構成する系列データより特徴量の生成方法が推定され、推定された生成方法で、前記系列データより特徴量が生成される。

図面の簡単な説明

- [0011] [図1]本開示のフローデータを説明する図である。
- [図2]フローデータにおけるセッション単位、時刻単位、属性データ、および時系列データの例を説明する図である。
- [図3]本開示の情報処理装置の構成例を説明するハードウェアブロック図である。
- [図4]図3のUI制御部、データ処理部、および機械学習モデル生成部により実現される機能を説明する機能ブロック図である。
- [図5]フローデータにおける属性データおよび時系列データの構成例を説明する図である。
- [図6]フローデータにおけるセッション単位の列、時刻単位の列、および予測対象列の設定を促すUIの表示画像例を説明する図である。

[図7]出力フォーマットとしてのメルトフォーマットの例を説明する図である。

[図8]出力フォーマットとしてのピボットフォーマットの例を説明する図である。

[図9]野球の所定の打者の投球ログに係るフローデータに基づいて、球速列を予測対象に設定した場合の出力フォーマットの例を説明する図である。

[図10]野球の所定の打者の投球ログに係るフローデータに基づいて、結果列を予測対象に設定した場合の出力フォーマットの例を説明する図である。

[図11]時系列データの特徴量の生成方法を説明する図である。

[図12]特徴量の生成に係る窓の設定例を説明する図である。

[図13]特徴量の生成に係る窓のその他の設定例を説明する図である。

[図14]特徴量の生成元となる系列データの選択を説明する図である。

[図15]セッション内特徴量データの生成例を説明する図である。

[図16]セッション間特徴量データの生成例を説明する図である。

[図17]本開示のフローデータにおけるセッションID、時刻単位、属性データ、時系列データ、セッション内特徴量、セッション間特徴量、およびセッション集合IDのそれぞれの例を説明する図である。

[図18]特徴量データの提示例を説明する図である。

[図19]特徴データ生成処理を説明するフローチャートである。

[図20]生成元選択処理を説明するフローチャートである。

[図21]セッション内特徴量データ生成処理を説明するフローチャートである。

[図22]セッションをクラスタリングする変形例を説明する図である。

[図23]セッションをクラスタリングする変形例を説明する図である。

[図24]汎用のコンピュータの構成例を示している。

発明を実施するための形態

[0012] 以下に添付図面を参照しながら、本開示の好適な実施の形態について詳細に説明する。なお、本明細書及び図面において、実質的に同一の機能構成を

有する構成要素については、同一の符号を付することにより重複説明を省略する。

[0013] 以下、本技術を実施するための形態について説明する。説明は以下の順序で行う。

1. 本開示の概要
2. 好適な実施の形態
3. ソフトウェアにより実行させる例

[0014] <<1. 本開示の概要>>

<フローデータ>

本開示は、特に、時系列データから機械学習モデルの作成に有効な特徴量を効率的に探索して抽出できるようにするものである。

[0015] 本明細書においては、複数の時系列データからなるデータ集合をフローデータと称するものとし、フローデータから機械学習モデルの作成に有効な特徴量を効率的に探索して抽出する技術を説明する。

[0016] そこで、まず、本明細書において使用する用語について定義する。

[0017] フローデータは、1つ以上の時系列データを必須構成とし、オプションとして1つ以上の属性データを備えることができるデータセットである。すなわち、フローデータにおいては、少なくとも1つ以上の時系列データが必ず含まれるのに対して、属性データは含まれないことがあってもよいが、複数の属性データが含まれてもよい。

[0018] ここで、時系列データは、時間の経過に伴って変化するデータであり、属性データは時間の経過に対し不変なデータである。

[0019] 例えば、病院内の患者につけたセンサから取得されるバイタル信号の場合、例えば、心拍や単位時間あたりの呼吸数、および測定装置における動作ログなどが、時系列データとなり、測定対象となる各患者の性別や体重などが属性データとなる。

[0020] これらの時系列データと属性データとが患者毎に蓄積されるとき、一人の患者を集合単位とするフローデータが構成される。

- [0021] また、工場で使用されるロボットアームの稼働状態をセンサで計測する場合、ロボットアームから取得できるセンサデータが時系列データとなり、個体別の故障回数などが属性データとなる。
- [0022] そして、これらの時系列データと属性データとがロボットアーム毎に蓄積されるとき、1つのロボットアームを集合単位とするフローデータが構成される。
- [0023] さらに、野球の試合の投球履歴が蓄積される場合、打席内で投球された球の速度等が、時系列データとなり、投手や打者の情報が属性データとなる。
- [0024] そして、これらの時系列データと属性データとが打席毎に蓄積されるとき、1つの打席を集合単位とするフローデータが構成される。
- [0025] すなわち、図1で示されるように、フローデータは、矢印で示される時間軸上の丸印で示されるタイミングにおいて時系列に測定される、データ $D_t 1$ 、 $D_t 2$ などからなる時系列データと、測定対象の人物の性別や体重などのデータ $D_a 1$ や、測定装置の装置名や、測定装置の設定値などのデータ $D_a 2$ などからなる属性データとから構成される。
- [0026] 尚、図1においては、病院において、患者のバイタル信号からなるデータ $D_t 1$ 、 $D_t 2$ を時系列データとし、患者の性別や体重のデータ $D_a 1$ 、および測定装置の装置名および設定値のデータ $D_a 2$ を属性データとした時の例である。
- [0027] また、時系列データを構成するデータ $D_t 1$ で示されるように、丸印で示される個々の時系列データの時間間隔は、間隔 T_1 、 T_2 で示されるように不均等でもよいし、図示しないが均等であってもよい。
- [0028] さらに、時系列データと属性データとからなるフローデータが、例えば、患者毎、測定装置毎、設定値毎等の一つの集合を構成するとき、この集合単位をセッションと称する。図1においては、所定の条件で構成されるフローデータの集合がセッション SS であることが示されている。
- [0029] そして、複数のセッション SS からなるフローデータに基づいて、様々な予測対象が予測される。

[0030] 図2は、病院バイタルログ、工場ロボットログ、および野球投球ログのそれぞれがフローデータを構成する場合の、セッション単位の例、時刻単位の例、属性データの例、および時系列データの例をまとめたものである。

[0031] すなわち、フローデータが病院バイタルログから構成されるものである場合、セッション単位の例は、患者であり、時刻単位の例は、日時であり、属性データの例は、患者の性別であり、時系列データの例は、心拍信号である。

[0032] また、フローデータが工場ロボットログである場合、セッション単位の例は、ロボットであり、時刻単位の例は、日時であり、属性データの例は、ロボットの故障回数であり、時系列データの例は、トルクセンサ信号である。

[0033] さらに、フローデータが野球投球ログである場合、セッション単位の例は、打席であり、時刻単位の例は、打席内球数であり、属性データの例は、投手の左／右投げであり、時系列データの例は、球速である。

[0034] この様にフローデータは、様々な実体として存在し、IoTの普及に伴って今後大量に生成されうるデータである。

[0035] ところで、フローデータに対して機械学習モデルによる予測が行われる場合、フローデータから機械学習モデル用の特徴量を作成する必要があるが、予測精度に寄与する機械学習モデル用の特徴量を作成する処理（特徴量エンジニアリング）は、手間と時間が掛かる処理であった。

[0036] より具体的には、一般ユーザは自分が予測したい対象系列や時刻情報については理解しているものの、ユーザが行いたいタスクに対する機械学習モデルを構築する上で必要なデータ加工については加工方法がわからないことが多い。

[0037] また、機械学習モデル用の特徴量を生成するツールがいくつか提案されているが、時系列データが等間隔である・時系列データの将来値予測のみに対応するなど、予測対象に対する制約が存在し、ユーザが行いたい予測対象の予測を網羅できないことが多い。

[0038] さらに、フローデータは膨大かつ複数時系列に渡ることが多いため、ユー

ザのデータセットの関連性への理解に限界があり、系列同士の関連性を踏まえてデータセットから特徴量を作成することが困難または煩雑になる。

[0039] また、反対に、全く事前知識なしに総当たりで特徴量を作成すると無駄な特徴量を作成することになり、不要な計算コストがかかってしまう。

[0040] そこで、本開示においては、ユーザにより、フローデータに対して、最低限の設定入力がなされることで、幅広いタスクに対して有効な特徴量を容易に生成できるようにする。

[0041] より具体的には、本開示においては、フローデータにおける、時刻を示す列およびセッション単位を示す列、並びに予測対象列が、ユーザにより入力されると、時系列データの将来値の予測、時系列データにおいて特定のイベントが発生するか否かの予測、および、時系列でない（時刻によって変化しない）データの予測等に有効な特徴量を、現実的な時間内で生成することが可能となる。

[0042] <<2. 好適な実施の形態>>

<本開示の情報処理装置の構成例>

次に、図3を参照して、本開示の情報処理装置の構成例について説明する。

[0043] 情報処理装置31は、制御部51、入力部52、出力部53、記憶部54、通信部55、ドライブ56、およびリムーバブル記憶媒体57より構成されており、相互にバス58を介して接続されており、データやプログラムを送受信することができる。

[0044] 制御部51は、プロセッサやメモリから構成されており、情報処理装置31の動作の全体を制御する。また、制御部51は、UI制御部61、データ処理部62、および機械学習モデル生成部63を備えている。

[0045] UI制御部61は、フローデータの入力を受け付けると、タスク設定としての時刻を示す列やセッション単位を示す列、および、予測対象となる列の入力を促すUI (User Interface) を生成し、出力部53の表示部71や音声出力部72を制御して提示する。

- [0046] そして、UI制御部61は、これに応じてユーザにより入力部52が操作されることで、入力されるタスク設定を受け付けて、入力されたフローデータと共にデータ処理部62に出力する。
- [0047] また、UI制御部61は、出力部53の表示部71や音声出力部72を制御して、データ処理部62により生成された特徴量の情報を、出力部53の表示部71や音声出力部72を制御してユーザに提示する。
- [0048] データ処理部62は、UI制御部61より供給されるフローデータとタスク設定を取得して、機械学習モデルの生成において有効な特徴量（以下、有効特徴量とも称する）を特徴量データとして生成し、UI制御部61、および機械学習モデル生成部63に出力する。
- [0049] 機械学習モデル生成部63は、データ処理部62より供給される有効特徴量からなる特徴量データに基づいて、機械学習モデルを生成する。
- [0050] 尚、UI制御部61、およびデータ処理部62により実現される機能の詳細については、図4の機能ブロック図を参照して後述する。
- [0051] 入力部52は、ユーザが操作コマンドを入力するキーボード、マウス、タッチパネルなどの入力デバイスより構成され、入力された各種の信号を制御部51に供給する。
- [0052] 出力部53は、制御部51により制御され、表示部、および音声出力部を備えている。出力部53は、操作画面や処理結果の画像を、LCD (Liquid Crystal Display) や有機EL (Electro Luminescence) などからなる表示デバイスからなる表示部71に出力して表示する。また、出力部53は、音声出力デバイスからなる音声出力部72を制御して、各種の音声や音楽、効果音などを再生する。
- [0053] 記憶部54は、HDD (Hard Disk Drive) 、SSD (Solid State Drive) 、または、半導体メモリなどからなり、制御部51により制御され、各種のデータおよびプログラムを書き込む、または、読み出す。
- [0054] 通信部55は、制御部51により制御され、有線または無線により、LAN (Local Area Network) やブルートゥース (登録商標) 等に代表される通信を

実現し、必要に応じてネットワークを介して、各種の装置との間で各種のデータやプログラムを送受信する。

[0055] ドライブ56は、磁気ディスク（フレキシブルディスクを含む）、光ディスク（CD-ROM（Compact Disc-Read Only Memory）、DVD（Digital Versatile Disc）を含む）、光磁気ディスク（MD（Mini Disc）を含む）、もしくは半導体メモリなどのリムーバブル記憶媒体57に対してデータを読み書きする。

[0056] <UI制御部およびデータ処理部により実現される機能>

次に、図4の機能ブロック図を参照して、UI制御部61およびデータ処理部62により実現される機能について説明する。

[0057] UI制御部61は、フローデータ入力部101、タスク設定部102、および生成特微量可視化部103を備えている。

[0058] フローデータ入力部101は、入力部52の操作入力、記憶部54、通信部55、およびドライブ56を介したリムーバブル記憶媒体57の少なくともいずれかからフローデータの入力を受け付けて、データ処理部62、および生成特微量可視化部103に出力する。

[0059] タスク設定部102は、データ処理部62より供給されるフローデータの列推定結果を取得すると、列推定結果を示すUIを生成すると共に、その列推定結果を提示するUI上にタスク設定としての時刻を示す列やセッション単位を示す列、および、予測対象の列の入力を促す情報を付加して、出力部53の表示部71や音声出力部72を制御して提示する。タスク設定部102は、さらに、予測対象列の予測頻度と予測時刻についてもタスク設定として入力を促すようにしてもよい。

[0060] タスク設定部102は、このUIに促されてユーザにより入力部52が操作されて、時刻を示す列やセッション単位を示す列、および、予測対象、並びに、必要に応じて、予測対象列の予測頻度と予測時刻を加えたタスク設定の情報をデータ処理部62に出力する。

[0061] 尚、タスク設定については、図5、図6を参照して詳細を後述する。

- [0062] 生成特微量可視化部103は、フローデータ入力部101より供給されるフローデータと、データ処理部62より供給される有効特微量からなる特微量データとを取得するとUIとして可視化し、出力部53の表示部71や音声出力部72を制御して提示する。
- [0063] 尚、生成特微量可視化部103により特微量データの提示例については、図13を参照して、詳細を後述する。
- [0064] データ処理部62は、列推定部121、出力フォーマット決定部122、生成元選択部123、セッション内特微量生成部124、特微量選択部125、セッション間特微量生成部126、結合部127、特微量データストレージ128、およびループ判定部129を備えている。
- [0065] 列推定部121は、UI制御部61より供給されるフローデータのデータフォーマットを解析し、時刻を示す列やセッション単位を示す列などとなり得る列を推定し、列推定結果としてUI制御部61に出力する。
- [0066] 出力フォーマット決定部122は、UI制御部61のタスク設定部102より供給されるタスク設定としての時刻を示す列やセッション単位を示す列、および、予測対象の列の情報に基づいて、フローデータの出力フォーマットを決定し、生成元選択部123に出力する。
- [0067] この際、出力フォーマット決定部122は、タスク設定として予測対象列の予測頻度と予測時刻についても供給される場合、予測対象列の予測頻度と予測時刻についても考慮した出力フォーマットを決定する。
- [0068] 尚、出力フォーマットの決定については、図7乃至図10を参照して、詳細を後述する。
- [0069] 生成元選択部123は、出力フォーマット決定部122より供給される出力フォーマットに従って、フローデータから、特微量の生成元となる系列データを選択的に抽出する処理を実行して、処理結果をセッション内特微量生成部124に出力する。
- [0070] 尚、特微量の生成元となる系列データを選択的に抽出する処理については、図14を参照して、詳細を後述する。

- [0071] セッション内特徴量生成部 1 2 4 は、生成元選択部 1 2 3 より供給されるフローデータのうち、特徴量の生成に必要とされる系列データに基づいて、セッション内特徴量データを生成して特徴量選択部 1 2 5 に出力する。
- [0072] より詳細には、セッション内特徴量生成部 1 2 4 は、メタデータ抽出部 1 2 4 a、および推定モデル 1 2 4 b を備えている。
- [0073] メタデータ抽出部 1 2 4 a は、フローデータ内の時系列データの数や系列長（1 系列あたりのサンプル数）や、各変数の統計値（平均および分散等）などからなるメタデータを抽出し、推定モデル 1 2 4 b に出力する。
- [0074] 推定モデル 1 2 4 b は、メタデータと、予測対象の予測に用いる特徴量の生成方法とをペアにして予め学習されたモデルであり、メタデータに基づいて、予測対象の予測に必要とされる特徴量の生成方法を推定する。
- [0075] セッション内特徴量生成部 1 2 4 は、推定モデル 1 2 4 b によりメタデータから推定された特徴量の生成方法により、生成元選択部 1 2 3 により選択された生成元となる系列データを用いてセッション内特徴量データを生成する。
- [0076] 尚、セッション内特徴量データの生成方法については、図 1 5 を参照して、詳細を後述する。
- [0077] 特徴量選択部 1 2 5 は、セッション内特徴量データやセッション間特徴量データを構成する特徴量のそれぞれについて、予測対象に対する予測に係る有効度スコアを求めて、所定の有効度スコアよりも高い特徴量を選択し、それ以外を除外して、セッション内特徴量データやセッション間特徴量データを再構成する。
- [0078] より詳細には、特徴量選択部 1 2 5 は、セッション内特徴量選択部 1 4 1、セッション間特徴量選択部 1 4 2、および有効度スコア算出部 1 4 3 を備えている。
- [0079] セッション内特徴量選択部 1 4 1 は、有効度スコア算出部 1 4 3 を制御して、セッション内特徴量データを構成する特徴量のそれぞれについて有効度スコアを算出させ、所定の有効度スコアよりも高い特徴量を選択すると共に

、所定の有効度スコアよりも低い特徴量を除外することで、セッション内特徴量データを再構成して、セッション間特徴量生成部126、および結合部127に出力する。

[0080] セッション間特徴量選択部142は、有効度スコア算出部143を制御して、セッション間特徴量生成部126より供給されるセッション間特徴量データを構成する特徴量のそれぞれについて有効度スコアを算出させ、所定の有効度スコアよりも高い特徴量を選択して、所定の有効度スコアよりも低い特徴量を除外することで、セッション間特徴量データを再構成して、結合部127に出力する。

[0081] 有効度スコア算出部143は、セッション内特徴量データ、およびセッション間特徴量データを構成する特徴量のそれぞれについて、例えば、予測対象との相互情報量を有効度スコアとして算出し、セッション内特徴量選択部141およびセッション間特徴量選択部142、並びにループ判定部129に出力する。

[0082] また、有効度スコア算出部143は、セッション内特徴量データ、およびセッション間特徴量データを用いて生成される、機械学習モデルに対する精度を有効度スコアとして求めるようにしてもよい。ただし、この場合、機械学習モデルは、機械学習モデル生成部63により生成される機械学習モデルよりも簡易的な機械学習アルゴリズム又はハイパパラメタにより求められる機械学習モデルを用いるものとする。

[0083] この場合、セッション内特徴量選択部141、およびセッション間特徴量選択部142は、生成された機械学習モデルの精度等から計算された有効度スコアが所定値よりも低下しない、セッション内特徴量データやセッション間特徴量データのそれぞれを構成する特徴量の部分集合を選択して、セッション内特徴量データ、およびセッション間特徴量データを再構成する。

[0084] セッション間特徴量生成部126は、特徴量選択部125より出力される、所定の有効度スコアよりも高い特徴量からなる再構成されたセッション内特徴量データに基づいて、セッション間特徴量データを生成して、特徴量選

択部 125 に出力する。

- [0085] 尚、セッション間特徴量データの生成方法については、図 16 を参照して、詳細を後述する。
- [0086] 結合部 127 は、特徴量選択部 125 より供給される所定の有効度スコアよりも高い特徴量からなる再構成されたセッション内特徴量データと、セッション間特徴量データとを結合して、特徴量データを構成し、特徴量データストレージ 128 に格納する。
- [0087] 特徴量データストレージ 128 は、結合部 127 より供給される特徴量データをストレージすると共に、必要に応じてループ判定部 129 に対して、ストレージしている特徴量データを供給する。
- [0088] ループ判定部 129 は、特徴量データストレージ 128 に格納されたセッション内特徴量データと、セッション間特徴量データとが結合された特徴量データを構成する特徴量の有効度スコアに基づいて、予測対象を予測する上での特徴量データの全体としての有効度スコアを、例えば、全体の平均値として算出する。
- [0089] ループ判定部 129 は、特徴量データの全体としての有効度スコアが所定値よりも低い場合、再度生成元選択部 123 に同一のフローデータから、現状の特徴量の数よりも多くの特徴量を抽出するように、処理を再度ループさせるように指示する。
- [0090] そして、ループ判定部 129 は、所定の時間が経過したとき、または、有効度が所定値よりも高い場合、その時点で特徴量データストレージ 128 に格納された特徴量データと、特徴量データの全体の有効度スコアの情報を UI 制御部 61、および機械学習モデル生成部 63 に出力する。
- [0091] 生成特徴量可視化部 103 は、生成された特徴量データと、特徴量データの全体の有効度スコアの情報を、UI として可視化して提示する。
- [0092] この際、例えば、生成された特徴量データの有効度スコアが十分であるとみなされて、ユーザにより、選択された特徴量データに基づいた、機械学習モデルの生成が指示されるとき、機械学習モデル生成部 63 は、供給された

特徴量データに基づいて、機械学習モデルを生成するようにしてもよい。

[0093] <タスク設定について>

タスク設定は、フローデータから、時系列データの将来値の予測、時系列データにおいて特定のイベントが発生するか否かの予測、および時系列でない（時刻によって変化しない）データの予測といったタスクを実現するための設定である。

[0094] より具体的には、タスク設定は、フローデータにおける時刻を示す列、またはセッション単位を示す列、および予測対象となる列の設定であり、さらに、必要に応じて、予測対象列の予測頻度と予測時刻の設定も含む。

[0095] 例えば、図5で示されるようなフローデータである場合におけるタスク設定について説明する。

[0096] 図5は、野球の所定の打者の投球ログに係るフローデータの例を示している。図5のフローデータFDは、属性データADおよび時系列データTDから構成されている。

[0097] 属性データADは、3つのデータ列から構成されており、図中の左から投手ID列、打席ID列、および結果列から構成される。

[0098] 投手ID列は、所定の打者に対して投球した投手を識別するIDが登録される列であり、図中においては、上から投手ID=A, B, Aが登録されている。

[0099] 打席ID列は、所定の打者の打席を識別するIDが登録される列であり、図中においては、上から打席ID=0, 1, 2が登録されている。

[0100] 結果列は、所定の打者の、投手IDで識別される投手による投球に対する、打席IDで識別される打席における結果が登録される列であり、図中においては、上から「ヒット」、「アウト」、および「アウト」が登録されている。

[0101] これにより、所定の打者は、投手ID=Aの投手による投球に対して、打席ID=0で識別される打席において、ヒットを放ったことが登録されている。

[0102] また、所定の打者は、投手ID=Bの投手による投球に対して、打席ID=1で識別される打席において、アウトになったことが登録されている。

[0103] さらに、所定の打者は、投手ID=Aの投手による投球に対して、打席ID=

- 2で識別される打席において、アウトになったことが登録されている。
- [0104] 時系列データTDは、3つのデータ列から構成されており、図中の左から打席ID列、投球ID列、および球速列から構成される。
- [0105] 打席ID列は、所定の打者の打席を識別するIDが登録される列であり、図中においては、上から打席ID=0, 0, 0, 1, 1, 2, 2, 2が登録されている。
- [0106] 投球ID列は、所定の打者に対する、投手から投げられた投球を識別するIDが、時系列に登録される列であり、図中においては、上から投球ID=0, 1, 2, 0, 1, 0, 1, 2が登録されている。
- [0107] 球速列は、所定の打者が投手IDで識別される投手により、打席IDで識別される打席において投げられた球速(km/h)が登録される列であり、図中においては、上から140, 150, 120, 120, 110, 90, 130, 155が登録されている。
- [0108] これにより、所定の打者の、打席ID=0で識別される打席において、投球ID=0で識別される最初の投球の球速が140km/hであり、投球ID=1で識別される2球目の投球の球速が150km/hであり、投球ID=2で識別される3球目の球速が120km/hであることが登録されている。
- [0109] また、所定の打者の、打席ID=1で識別される打席において、投球ID=0で識別される最初の投球の球速が120km/hであり、投球ID=1で識別される2球目の投球の球速が110km/hであることが登録されている。
- [0110] さらに、所定の打者の、打席ID=2で識別される打席において、投球ID=0で識別される最初の投球の球速が90km/hであり、投球ID=1で識別される2球目の投球の球速が130km/hであり、投球ID=2で識別される3球目の投球の球速が155km/hであることが登録されている。
- [0111] この場合、時系列データTDにおける投球列の情報は、時系列に登録される情報であるため、時刻列とされる。
- [0112] また、時系列データTDと、属性データADとのそれぞれには、セッション列として、共通の打席ID列が存在する。

- [0113] さらに、投手ID列は、セッション列としての打席ID列の上位としてクラスタリングされた集合（セッションクラスタ）であるとも考えることもできる。
- [0114] 尚、時刻列は順序のわかる値（float, int）、または日時型（YY:MM:DD hh:mm:ssなど）であればよい。
- [0115] データ処理部62の列推定部121は、例えば、図5で示されるような時刻列やセッション列を推定して列推定結果として、UI制御部61のタスク設定部102に供給する。
- [0116] このUIに基づいて、入力部52が操作されて、時刻を示す列およびセッション単位を示す列、予測対象列、並びに、予測対象列の予測頻度と予測時刻を設定する情報が入力されると、タスク設定部102は、データ処理部の出力フォーマット決定部122に出力する。
- [0117] より詳細には、タスク設定部102は、列推定結果に基づいて、出力部53の表示部71や音声出力部72を制御して、フローデータをユーザに提示する。
- [0118] この際、タスク設定部102は、タスク設定として、時刻単位を示す列およびセッション単位を示す列、および予測対象列を設定するように促すUIを提示して、UIに応じて設定されたタスク設定の情報をデータ処理部62の出力フォーマット決定部122に出力する。
- [0119] より具体的には、タスク設定部102は、例えば、図6で示されるようなUIからなる表示画像PVを提示する。
- [0120] 図6の表示画像PVにおいて提示されるUIにおいては、上段に「時刻単位を示す列およびセッション単位を示す列、並びに予測対象列を設定して下さい。」と表記されており、時刻単位を示す列およびセッション単位を示す列、並びに予測対象列の設定を促す情報が提示されている。
- [0121] また、その下には、左側に属性データADが表示され、右側に時系列データTDが表示されている。
- [0122] さらに、設定を促す情報の提示に応じて、図6においては、点線で示される打席ID列が、セッション単位を示す列として設定され、一点鎖線で示され

る投球列が、時刻単位を示す列として設定され、実線で示される球速列が、予測対象として設定されている例が示されている。

[0123] タスク設定部102は、図6で示される点線、一点鎖線、および実線で示されるような枠などを用いて設定された、時刻単位を示す列およびセッション単位を示す列、および予測対象列の情報を出力フォーマット決定部122に出力する。

[0124] 尚、この際、時刻単位を示す列およびセッション単位を示す列、および予測対象列の情報に加えて、予測対象列の予測頻度と予測時刻をタスク設定として入力させるようにしてもよい。

[0125] <出力フォーマットの決定>

出力フォーマット決定部122は、UI制御部61のタスク設定部102より供給される時刻を示す列およびセッション単位を示す列、予測対象列、並びに、予測対象列の予測頻度と予測時刻を設定する情報に基づいて、出力フォーマットを決定する。

[0126] 出力フォーマットの形式は、例えば、図7で示されるようなメルトフォーマットや、図8のピボットフォーマットなどである。

[0127] 図7のメルトフォーマットは、左から、id列、time列、name列、およびvalue列から構成される。図7のメルトフォーマットでは、name列が、セッション単位を構成し、id列がセッション単位をグルーピングするセッションクラスタであり、time列がサンプリングの時刻列であり、value列がサンプリングされた時系列データ列となる。

[0128] すなわち、図7においては、セッション単位をグルーピングする上位のセッションクラスタについて、A、Bの2つ存在し、セッション単位内においてx、yの2つの系列があり、それぞれの系列において時刻t1、t2が設定されている。

[0129] このようにセッション単位で系列毎に、時刻設定が異なる場合、図7で示されるようなメルトフォーマットが有効である。

[0130] 図7においては、時系列データとして、上からx(A, t1), x(A,

t 2), y (A, t 1), y (A, t 2), x (B, t 1), x (B, t 2), y (B, t 1), y (B, t 2) が登録されている。

[0131] これに対して、セッション単位で全ての系列について、サンプリングの時刻列が共通である場合、図8で示されるようなピボットフォーマットでもよい。

[0132] すなわち、図8においては、id列、time列、value x列、およびvalue y列から構成される。図8のピボットフォーマットにおいては、セッション単位でx, yの2つの系列において、サンプリングの時刻列が共通化されており、value x列、およびvalue y列が、並列に登録される。

[0133] 図8においては、value x列として、上からx (A, t 1), x (A, t 2), x (B, t 1), x (B, t 2) が登録され、value y列として、y (A, t 1), y (A, t 2), y (B, t 1), y (B, t 2) が登録されている。

[0134] より具体的には、例えば、図9の左部で示されるように、点線で示される打席ID列がセッション単位を示す列として設定され、一点鎖線で示される投球列が、時刻単位を示す列として設定され、実線で示される球速列が、予測対象列として設定されている場合、出力フォーマット決定部122は、例えば、図9の右部で示されるような出力フォーマットF1S1に決定する。

[0135] 図9の右部における出力フォーマットF1S1は、図7を参照して説明したメルトフォーマットから構成されており、左からセッション単位を示す打席ID列、投球列、球速列、1球前球速列、および前打席結果列が設けられている。

[0136] 図9の右部における出力フォーマットF1S1において、打席ID列には、上から0, 0, 0, 1, 1, 2, 2, 2が登録され、投球列には、上から0, 1, 2, 0, 1, 0, 1, 2が登録されている。

[0137] また、球速列には、上から140, 150, 120, 120, 110, 90, 130, 155が登録され、1球前球速列には、上からNaN, 140, 150, NaN, 120, NaN, 90, 130が登録され、前打席結果列に上からN

aN, NaN, NaN, ヒット, ヒット, アウト, アウト, アウトが登録されている。

[0138] すなわち、ここでは、予測対象となる列が「球速」であるので、球速列のデータが、時系列データとして配置されるように、1時刻1行（1投球1列）のフォーマットとされる。

[0139] また、例えば、図10の左部で示されるように、点線で示される打席ID列がセッション単位を示す列として設定され、一点鎖線で示される投球列が、時刻単位を示す列として設定され、実線で示される結果列が、予測対象列として設定されている場合、出力フォーマット決定部122は、例えば、図10の右部で示されるような出力フォーマットF1S2に決定する。

[0140] 図10の右部における出力フォーマットF1S2は、図8を参照して説明したピボットフォーマットから構成されており、左から投手ID列、セッション単位を示す打席ID列、結果列、各打席球速平均列、および前打席結果列が設けられている。

[0141] 図10の右部における出力フォーマットF1S2において、投手ID列には、上からA, B, Aが登録され、打席ID列には、上から0, 1, 2が登録され、結果列に上からヒット、アウト、アウトが登録され、各打席球速平均列には、上から145, 115, 110が登録され、前打席結果列に上からNaN, ヒット, アウトが登録されている。

[0142] すなわち、ここでは、予測対象が「結果」であるので、1行は1セッション（1行1打席ID列）となる。時系列データは、各打席球速平均とされ、統計量を使って時間情報が集約された特徴量が追加された形式とされている。

[0143] <特徴量の生成方法>

次に、系列データ毎の特徴量の生成方法について説明する。

[0144] 特徴量は、時系列データから時間方向に得られる系列データ毎の複数の統計量を要素とするベクトルとして構成される。

[0145] 例えば、所定の系列の時系列データが、図11で示されるように、時間t方向に変化する波形L d tで表現される場合、時刻t1, t2, t3...

のそれぞれにおける、時間方向に所定の時間幅 w の窓を設定し、各窓の波形 $L d t$ の値が、部分系列 $X 1, X 2, X 3, \dots$ として取得される。

[0146] また、部分系列 $X 1, X 2, X 3, \dots$ のそれぞれに対応する所定時刻だけ未来の時刻 $t 1 1, t 1 2, t 1 3$ における波形 $L d t$ の値が、予測対象 $y 1, y 2, y 3, \dots$ として取得される。

[0147] そして、取得された部分系列 $X 1, X 2, X 3, \dots$ については、それぞれ時刻 $t 1, t 2, t 3, \dots$ から時間方向に所定時間幅 w の窓における所定の統計値 $f (X 1), f (X 2), f (X 3), \dots$ に変換され、変換された統計値と、予測対象 $y 1, y 2, y 3, \dots$ とを要素とするベクトルが構成されることにより、系列データ毎の特徴量が形成される。

[0148] より詳細には、所定の系列データの特徴量からなる列および予測対象列は、例えば、以下の式 (1)、式 (2) のように表現される。

$$[0149] \quad F s = (f (X 1), f (X 2), f (X 3), \dots) \quad \dots (1)$$

$$F p = (y 1, y 2, y 3, \dots) \quad \dots (2)$$

[0150] ここで、 $F s$ は、所定の系列データの特徴量であり、 $f (X 1), f (X 2), f (X 3), \dots$ は、それぞれ波形 $L d t$ で表現される所定の系列データの部分系列 $X n$ の統計量からなる要素である。また、 $y 1, y 2, y 3, \dots$ は、それぞれ部分系列 $X 1, X 2, X 3, \dots$ に対応する予測対象である。

[0151] また、部分系列 $X n$ に対応する系列データの特徴量 $F s$ を構成する各要素 $f (X n)$ は、例えば、以下の式 (3) のように表現される。

$$[0152] \quad f (X n) = (A v e (X n), M i n (X n), M a x (X n), V a r (X n), S t d e (X n), \dots) \quad \dots (3)$$

[0153] ここで、 $f (X n)$ は、部分系列 $X n$ の系列データの特徴量 $F s$ の各要素であり、 $A v e (X n)$ は、部分系列 $X n$ の平均値であり、 $M i n (X n)$

は、部分系列 X_n の最小値であり、 $\text{Max}(X_n)$ は、部分系列 X_n の最大値であり、 $\text{Var}(X_n)$ は、部分系列 X_n の分散であり、 $\text{Stde}(X_n)$ は、部分系列 X_n の標準偏差である。

[0154] 尚、部分系列 X_n に対しては、上述した平均値、最小値、最大値、分散、および標準偏差以外の他の統計量が用いられるようにしてもよい。

[0155] また、式 (3) の部分系列 X_n の特徴量 F_s の各要素 $f(X_n)$ は、各統計量を要素としたベクトルとして表現される例について説明してきたが、各統計量を用いたカーネル関数による重み付き積和（畳み込みカーネル）で表現されるようにしてもよい。尚、畳み込みカーネルについては、<https://arxiv.org/abs/1910.13051>等を参照されたい。

[0156] <窓の設定方法>

上述した部分系列 X_n を構成する窓は、様々な手法で設定されるようにしてもよい。

[0157] 例えば、図 12 の左部で示されるように、セッションの開始時刻 t_b を基準として、開始時刻からのオフセット offset_b を変化させながら、所定の時間幅 w_b で設定される窓 W_B を単位として部分系列 X_n が設定されてもよい。

[0158] また、図 12 の右部で示されるように、予測開始時刻を、例えば、基準時刻 t_s としたとき、基準時刻 t_s からのオフセット offset_s を変化させながら、所定の時間幅 w_s で設定される窓 W_S を単位として部分系列 X_n が設定されてもよい。

[0159] さらに、図 13 で示されるように、セッション開始時刻 t_b から予測実行時刻 t_s より所定時間オフセット (offset) された時刻 t_{os} までが時間幅で設定される窓 W_{SS} を単位として部分系列が設定されてもよい。

[0160] また、図 13 で示されるように、セッション開始時刻 t_b から終了までの全範囲が時間幅に設定される窓 W_A を単位として部分系列が設定されてもよい。

[0161] さらに、時刻 t_{os} から一定期間だけ前にshiftしたときの特定値 $Ldt(s)$ が

、部分系列として得られるようにしてもよい。

[0162] <特徴量の生成元となる系列の選択について>

次に、図14を参照して、生成元選択部123による特徴量の生成元となる系列データの選択について説明する。

[0163] 上述したように、特徴量の生成元となる情報（以降、生成元特徴量と称する）は、系列データ毎にベクトル化した情報として生成される。

[0164] しかしながら、全ての系列データが、予測対象の予測に有用な情報とは限らず、予測対象の予測には不要なものが存在する。

[0165] そこで、本開示においては、生成元選択部123が、フローデータから抽出される時系列データ、および属性データのいずれも含む系列データのうち、特徴量の生成元として有用である否かを判断して、必要に応じて除外する。

[0166] より具体的には、図14で示されるように、例えば、予測対象Tを予測する機械学習モデルの生成元に利用可能な、時系列データとして系列データL1乃至L3が存在している場合について考える。

[0167] 生成元選択部123は、系列データL1乃至L3のそれぞれが、予測対象の予測に用いる特徴量の生成元として適切であるか否かを判定する。より詳細には、例えば、生成元選択部123は、系列データL1について、予測対象Tの予測に用いる時系列の生成元特徴量として、統計量F a乃至F dからなる生成元特徴量F (t n)を時系列に抽出して、特徴量テーブルTBを生成する。

[0168] 尚、ここでいう、統計量F a乃至F dは、上述した式(3)のAve (X n), Min (X n), Max (X n), Var (X n), Stde (X n), …に対応し、生成元特徴量F (t n)は、系列データの特徴量F sを構成する各要素f (X n)に対応する。

[0169] 図14においては、系列データL1より抽出される生成元特徴量F (t n)として、生成元特徴量F (t 1) (= (F a (t 1), F b (t 1), F c (t 1), F d (t 1))), および生成元特徴量F (t 2) (= (F a (

t_2), $F_b(t_2)$, $F_c(t_2)$, $F_d(t_2)$), が抽出されて、特徴量テーブルTBが作成されている例が示されている。尚、図14の特徴量テーブルTBにおいて、予測対象Tの詳細な記載は省略されている。

[0170] 次に、生成元選択部123は、生成元特徴量 $F(t_1)$ ($= (F_a(t_1), F_b(t_1), F_c(t_1), F_d(t_1))$), $F(t_2)$ ($= (F_a(t_2), F_b(t_2), F_c(t_2), F_d(t_2))$)に基づいて、系列データL1が予測対象Tの予測に寄与する系列であるか否かを判定する。

[0171] まず、生成元選択部123は、例えば、系列データL1の時系列の変化がなく、また、予測対象との相関が認められない場合、系列データL1を特徴量から除外する。

[0172] そして、系列データL1が、時系列の変化があり、予測対象との相関があるものと認められた場合、生成元選択部123は、系列データL1の生成元特徴量 $F(t_1)$ ($= (F_a(t_1), F_b(t_1), F_c(t_1), F_d(t_1))$), $F(t_2)$ ($= (F_a(t_2), F_b(t_2), F_c(t_2), F_d(t_2))$), ...を、予測モデルPMに入力して、予測対象Tを予測し、その予測結果T'として求める。

[0173] 尚、予測モデルPMは、比較的簡易で、かつ、軽量の予測モデルであり、所定の系列の生成元特徴量に基づいて、簡易的に予測対象Tを予測するためのモデルである。

[0174] 生成元選択部123は、予測対象Tと予測結果T'との比較から予測精度PAを算出し、所定の閾値よりも予測精度が低い場合、系列データL1を、予測対象を求めるための生成元から除外する。

[0175] 生成元選択部123は、系列データL1のみならず、系列データL2, L3のそれぞれについても同様に予測精度PAを求めて、所定の予測精度よりも低い系列を生成元から除外する。

[0176] <セッション内特徴量データの生成方法>

次に、セッション内特徴量生成部124によるセッション内特徴量データの生成方法について説明する。

- [0177] セッション内特徴量生成部124は、図15で示されるように、メタデータ抽出部124aを制御して、フローデータに基づいて、例えば、フローデータにおける系列の本数、系列長、各系列の分散、および属性データの数などの情報を、フローデータのメタデータとして抽出させる。
- [0178] より詳細には、メタデータ抽出部124aは、フローデータのうちの、生成元選択部123において、生成元として選択された系列データからメタデータを抽出する。
- [0179] 尚、フローデータのメタデータは、フローデータに基づいて生成された特徴量データに基づいて機械学習モデル生成部63において生成される機械学習モデルやアルゴリズムなどでもよい。
- [0180] セッション内特徴量生成部124は、様々なメタデータと、最終的な有効特徴量の作成方法の分布とをペアの情報として取得してプールし、これらを使った学習により、メタデータに基づいて、有効特徴量の生成方法を推定する推定モデル124bを備えている。
- [0181] そこで、セッション内特徴量生成部124は、この推定モデル124bを制御して、抽出したフローデータのメタデータに基づいて、有効特徴量の生成方法を推定する。
- [0182] すなわち、推定モデル124bは、ユーザにより設定された時刻を示す列およびセッション単位を示す列、および予測対象列に基づいて、決定された出力フォーマットで抽出される系列データのうち、生成元選択部123により予測精度が所定の精度閾値よりも高い系列データから構成される、フローデータのメタデータに基づいて、有効特徴量の生成方法を推定する。
- [0183] これにより、フローデータを構成する系列データのうち、ユーザにより設定された時刻を示す列およびセッション単位を示す列、並びに予測対象列に設定された系列データのうち、予測対象の予測について、予測精度の高い系列データを用いた、有効特徴量の生成方法が推定されることになる。
- [0184] 結果として、ユーザにより設定された時刻を示す列およびセッション単位を示す列、並びに予測対象列を反映した、予測精度の高い機械学習モデルの

生成に最適な有効特徴量を生成することが可能となる。

[0185] 有効特徴量の生成方法の情報は、例えば、有効特徴量に使用する系列データの使用方法、窓の設定方法、および特徴量の要素における各値の割合や重みの設定方法など、有効特徴量の生成方法（＝算出方法）を特定する情報である。

[0186] より具体的には、有効特徴量の生成に使用する系列データの使用方法を特定する情報は、例えば、カテゴリ型の系列データと数値型の系列データの使用割合を、40：60などの所定の割合で使用するという情報である。

[0187] また、窓の設定方法を特定する情報は、例えば、図12の窓WB、窓WS、および図13の窓WSS、窓WAのそれぞれで得られた情報を、例えば、50：20：20：10で使用するという情報である。

[0188] さらに、特徴量の各要素における各値の割合や重みの設定方法を特定する情報は、例えば、 $Ave(X_n)$ 、 $Min(X_n)$ 、 $Max(X_n)$ 、 $Var(X_n)$ 、 $Stde(X_n)$ のそれぞれの割合や、重みを割り付けるといった情報である。

[0189] そして、セッション内特徴量生成部124は、各系列データを用いて、推定された作成方法で、有効特徴量を作成し、作成した有効特徴量を用いてセッション内特徴量データを生成し、出力する。

[0190] <セッション間特徴量データ>

次に、セッション間特徴量生成部126により生成されるセッション間特徴量データについて説明する。

[0191] セッション間特徴量生成部126は、以上のように、フローデータのうち、予測モデルPMに基づいて、予測精度PAが所定値よりも高い系列データを用いて、フローデータのメタデータから推定される有効特徴量の生成方法により生成された特徴量からなるセッション内特徴量データの時間の前後関係から得られる特徴量を用いて、セッション間特徴量データを生成する。

[0192] すなわち、例えば、フローデータが野球投球ログである場合、同一の打者、同一の投手、または、同一の打者で、かつ、同一の投手であるときのい

れかにおける、 x 打席前の特徴量、または、過去全体の特徴量は、セッション間特徴量として扱うことができる。

[0193] セッション単位として打席ID列が設定される場合、打席IDには順序が認められる、いわゆる整数型データであるので、セッション単位での前後関係を想定した時系列データとして扱い、セッション間特徴量とすることができる。

[0194] また、予測対象としてヒットやアウトなどの結果列が設定される場合、前打席のヒットまたはアウトといった情報は、いわゆるstring型データであり、順序が認められないが、時刻列の値に基づいて、順序が特定されるようにして、セッション間特徴量とすることができる。

[0195] さらに、セッション単位である打席ID列をクラスタリングする投手ID列の情報を集合として用いることでグルーピングされたセッション内のクラスを単位とした前後関係を計算するようにしてもよい。例えば、投手ID列を、セッション単位である打席IDをクラスタリングする集合とすることができる場合、“同じ投手の”前打席のヒット、またはアウトといった予測対象となる結果列や球速平均をセッション間特徴量とすることができる。

[0196] より具体的には、図16の左部で示されるように、セッションの単位として打席IDが設定され、時刻列として投球列が設定され、予測対象として結果列が設定されて、図16の中央部で示されるように、左から、セッションの単位である打席ID列をクラスタリングする投手ID列、セッションの単位である打席ID列、結果列、および各打席球速平均列からなるセッション内特徴量データが生成される場合について考える。

[0197] この場合、打席IDがセッション単位であるので、前打席の特徴量はセッション間特徴量である。このため、図16の中央部で示されるように、セッション内特徴量におけるセッション単位となる各打席球速平均列に対して、図16の右部で示されるように、セッション間特徴量においては、セッション単位となる打席ID列をクラスタリングする投手ID列に対応する投手の前回の球速平均列が追加されている。

[0198] 尚、図16の右部におけるセッション間特徴量における投手の前の球速平均列には、その右側の、図16の中央部におけるセッション内特徴量における各打席球速平均における投手IDがA、Bにおける打席IDが0、1のそれぞれの137km/h、および115km/hの値が、打席IDが2、3における同一投手の前の球速平均の値として記載されている。

[0199] また、投手IDがA、Bにおける打席IDが0、1の投手の前の球速平均の値は、存在しないので、「NaN」とされている。

[0200] さらに、セッション単位をクラスタリングするセッション集合を用いることで、同じセッション集合における前打席の特徴量なども作成できる。

[0201] <本開示のフローデータの例>

次に、図17を参照して、本開示のフローデータに対する、病院バイタルログ、工場ロボットログ、および野球投球ログのそれぞれがフローデータを構成する場合の、セッション（単位の）IDの例、時刻単位の例、属性データの例、時系列データの例、セッション内特徴量の例、セッション間特徴量の例、およびセッション単位をクラスタリングしたときのセッション集合（単位の）IDの例について説明する。

[0202] すなわち、フローデータが病院バイタルログである場合、セッション（単位の）IDの例は、患者IDであり、時刻単位の例は、日時であり、属性データの例は、患者の性別であり、時系列データの例は、心拍信号であり、セッション内特徴量の例は、患者の平均心拍であり、セッション間特徴量の例は、病院別患者の年齢であり、セッション集合（単位の）IDは、病院IDである。

[0203] また、フローデータが工場ロボットログである場合、セッション（単位の）IDの例は、稼働IDであり、時刻単位の例は、日時であり、属性データの例は、ロボットの設置場所であり、時系列データの例は、トルクセンサ信号であり、セッション内特徴量の例は、その日のロボットの平均停止回数であり、セッション間特徴量の例は、ロボット別通算停止回数であり、セッション集合（単位の）IDは、ロボットIDである。

[0204] さらに、フローデータが野球投球ログである場合、セッション（単位の）I

Dの例は、打席IDであり、時刻単位の例は、打席内球数であり、属性データの例は、投手の左／右投げであり、時系列データの例は、球速であり、セッション内特徴量の例は、打席内平均球速であり、セッション間特徴量の例は、同一投手の過去3打席の結果であり、セッション集合（単位の）IDは、投手IDである。

[0205] 尚、フローデータに対する、セッション（単位の）IDの例、時刻単位の例、属性データの例、時系列データの例、セッション内特徴量の例、セッション間特徴量の例、およびセッション単位をクラスタリングしたときのセッション集合（単位の）IDの例については、図17に限定されるものではない。

[0206] <特徴量データの提示例>

次に、図18を参照して、生成特徴量可視化部103によりフローデータと特徴量データとが可視化されて提示される際の提示例について説明する。

[0207] 図18は、フローデータが野球投球ログである場合の提示例を示している。

[0208] 図18の特徴量データの提示例においては、上段に特徴量データテーブルが表示され、下段には、上段の特徴量データテーブル内において、指定されたセッション内特徴量データの一部の詳細データを表示するグラフが表示されている。また、特徴量データテーブルの右上部には、特徴量データ全体の有効度スコア表示欄が設けられており、図18においては、「特徴量データ全体の有効度スコア：85／100」と表記されており、例えば、予測対象の予測に対する有効度スコアが100点満点中85点であることが示されている。

[0209] 特徴量データテーブルには、左からデータID列、セッション単位としての打席ID列、属性データとしての投手ID列、および結果列、時刻列としての投球ID列、予測対象としての球速列、セッション内特徴量データとしての1球前の球速列、直近3球球速平均列、および1球前の球種列、並びに、セッション間特徴量データとしての投手の前回の球速平均列が設けられている。

[0210] 図18においては、データID列においては、上から順に1, 2, 3, 4,

5, 6と表示されている。

[0211] また、セッション単位の列としての打席ID列においては、上から順に1, 1, 1, 1, 2と表示されており、データID=1乃至5までのデータが、打席ID=1のものであり、データID=6のデータが、打席ID=2のものであることが示されている。

[0212] さらに、属性データとしての投手ID列においては、上から順に、A, A, A, A, Bと表示されており、データID=1乃至5までのデータが、投手ID=Aのものであり、データID=6のデータが、投手ID=Bのものであることが示されている。

[0213] また、結果列においては、上から「ヒット」、「ヒット」、「ヒット」、「ヒット」、「ヒット」、「アウト」と表示されており、データID=1乃至5までの結果列が、ヒットであり、データID=6の結果列が、アウトであることが示されている。

[0214] 時刻列としての投球ID列においては、上から順に1, 2, 3, 4, 5, 1と表示されており、データID=1乃至5までの投球ID列が、打席ID=1における同一投手ID=Aの投手により投げられた第1球目から第5球目までのデータであり、投球ID=5の投球において打者がヒットを放ったことが示されている。

[0215] また、データID=6の投球ID=1の投球のデータであることが示されている。

[0216] 予測対象である球速列においては、上から順に143.9, 140.2, 130.9, 90.4, 124.3, 150.2と表示されており、打席ID=1の打席における、投球ID=1の1球目の球速が143.9km/hであり、投球ID=2の2球目の球速が140.2km/hであり、投球ID=3の3球目の球速が130.9km/hであり、投球ID=4の4球目の球速が90.4km/hであり、投球ID=5の5球目の球速が124.3km/hであり、打席ID=2の打席における、投球ID=1の1球目の球速が150.2km/hであることが示されている。

[0217] セッション内特徴量データとしての1球前の球速列においては、上からNaN

, 143.9, 140.2, 130.9, 90.4, NaNと表示されており、データID=1乃至6のそれぞれの1球前の球速が表示されている。

[0218] 直近3球球速平均列においては、上からNaN, NaN, NaN, 138.3, 120.5, NaNと表示されており、データID=1乃至6のそれぞれの直近3球の平均球速が表示されている。

[0219] 1球前の球種列においては、上からNaN、ストレート、スライダー、チェンジアップ、スローボール、NaNと表記されており、打席ID=1の打席において、投手ID=Aの投手が、投球ID=2乃至5の投球において1球前に投球された球種が、それぞれストレート、スライダー、チェンジアップ、スローボールであることが示されている。

[0220] セッション間特徴量データとしての投手の前の球速平均列においては、上から順に、120.4, 120.4, 120.4, 120.4, 144.2と表示されており、打席ID=1の打席の前の打席における、球速平均が120.4km/hであることが表示されており、打席ID=2の打席の前の打席における、球速平均が144.2km/hであることが表示されている。

[0221] さらに、下段においては、特徴量データテーブルのデータID=4である投球ID=4における、直近3球球速平均が指定された場合の、詳細データを表示するグラフ表示例が表示されている。

[0222] 下段のグラフにおいては、投球ID=1乃至5の球速が143.9km/h、140.2km/h、130.9km/h、90.4km/h、および124.3km/hであることを示す位置がプロットされ、プロットされた各点が直線により接続されたグラフが表示されている。

[0223] さらに、このうち、投球ID=4における、直近3球の球速が、それぞれ143.9km/h、140.2km/h、および130.9km/hであり、この球速平均が138.3km/hであることが表記されている。

[0224] 図18で示される例では、予測対象となる球速に対して、セッション内特徴量データとして、1球前の球速、直近3球球速平均、および1週前の球種が提示され、セッション間特徴量データとして、投手の前の球速平均が生

成されていることが提示される。

[0225] 図18で示されるような提示により、ユーザは、予測対象となる球速を予測する機械学習モデルを生成する上で、セッション内特徴量データとして、1球前の球速、直近3球球速平均、および1週前の球種が提案され、セッション間特徴量データとして、投手の前の球速平均が提案されたことを認識することができる。

[0226] また、有効度スコアが提示されることにより、特徴量データを用いて生成される機械学習モデルを用いた予測において期待される精度を、ある程度認識することが可能となる。

[0227] 結果として、フローデータを入力し、フローデータに対する時刻を示す列、およびセッション単位を示す列、並びに予測対象列を指定するだけで、機械学習モデルの生成に必要なとされる特徴量データを生成することが可能となる。

[0228] 尚、図18により提示される特徴量データを参照して、有効度スコアが低く、また、提示された特徴量データを参照しても、機械学習モデルの生成に十分な特徴量データが得られていないと判断される場合については、例えば、フローデータに対して指定する時刻を示す列、およびセッション単位を示す列をし直して、再度、特徴量データを生成させるようにしてもよいし、他のフローデータを使用するようにしてもよい。

[0229] <特徴量データ生成処理>

次に、図19のフローチャートを参照して、図4のUI制御部61、およびデータ処理部62の機能により実現される特徴量データ生成処理について説明する。

[0230] ステップS31において、フローデータ入力部101は、フローデータの入力を受け付けて、生成特徴量可視化部103、およびデータ処理部62に出力する。

[0231] ステップS32において、データ処理部62の列推定部121は、フローデータを解析して、フローデータを構成する列を推定し、推定結果をUI制御

部 6 1 に出力する。

[0232] ステップ S 3 3 において、タスク設定部 1 0 2 は、フローデータの列の推定結果を取得すると、推定結果と共に、タスク設定として、セッションの単位列、時刻の単位列、および予測対象の入力を促す、例えば、図 6 を参照して説明した表示画像 P V で示されるような UI を生成して、提示する。

[0233] そして、タスク設定部 1 0 2 は、ユーザからの入力を受け付けて、タスク設定として入力されたセッションの単位列、時刻の単位列、および予測対象の情報をデータ処理部 6 2 に出力する。

[0234] このとき、タスク設定部 1 0 2 は、タスク設定として、さらに、予測対象列の予測頻度と予測時刻の入力を促す情報も UI で提示し、予測対象列の予測頻度と予測時刻の情報についても入力を受け付けて、データ処理部 6 2 に出力する。

[0235] ステップ S 3 4 において、出力フォーマット決定部 1 2 2 は、タスク設定として供給されたセッションの単位列、時刻の単位列、および予測対象の情報に基づいて、フローデータより読み出す出力フォーマットを決定し、生成元選択部 1 2 3 に出力する。

[0236] ステップ S 3 5 において、生成元選択部 1 2 3 は、出力フォーマットに従って、フローデータより系列データを抽出すると共に、生成元選択処理を実行し、フローデータより出力フォーマットに基づいて抽出される系列データのうち、予測対象の予測に有効性の高い系列データを選択して、セッション内特徴量生成部 1 2 4 に出力する。

[0237] 尚、生成元選択処理については、図 2 0 のフローチャートを参照して、詳細を後述する。

[0238] ステップ S 3 6 において、セッション内特徴量生成部 1 2 4 は、セッション内特徴量生成処理を実行し、選択された系列データを用いて、セッション内特徴量データを生成して特徴量選択部 1 2 5 に出力する。

[0239] 尚、セッション内特徴量生成処理については、図 2 1 のフローチャートを参照して、詳細を後述する。

- [0240] ステップS 37において、特徴量選択部125のセッション内特徴量選択部141は、有効度スコア算出部143を制御して、供給されてきたセッション内特徴量データを構成する、それぞれの特徴量の予測対象の予測に係る有効度スコアを算出させると共に、算出した有効度スコアを自らとループ判定部129に出力させる。
- [0241] ステップS 38において、セッション内特徴量選択部141は、セッション内特徴量データを構成する、それぞれの特徴量のうち、有効度スコアが所定のスコア閾値よりも高い特徴量を有効特徴量として選択すると共に、その他の特徴量を除外し、有効特徴量からなるセッション内特徴量データを再構成してセッション間特徴量生成部126、および結合部127に出力する。
- [0242] ステップS 39において、セッション間特徴量生成部126は、特徴量選択部125より供給されたセッション内特徴量データを取得すると、記憶すると共に、他のセッション内特徴量データを利用して、セッション間特徴量データを生成し、特徴量選択部125に出力する。
- [0243] ステップS 40において、特徴量選択部125のセッション間特徴量選択部142は、有効度スコア算出部143を制御して、供給されてきたセッション間特徴量データを構成する、それぞれの特徴量の予測対象の予測に係る有効度スコアを算出させると共に、算出した有効度スコアを自らとループ判定部129に出力させる。
- [0244] ステップS 41において、セッション間特徴量選択部142は、セッション間特徴量データを構成する、それぞれの特徴量のうち、有効度スコアが所定のスコア閾値よりも高い特徴量を有効特徴量として選択すると共に、その他の特徴量を除外し、有効特徴量からなるセッション間特徴量データを再構成して結合部127に出力する。
- [0245] ステップS 42において、結合部127は、セッション内特徴量データと、セッション間特徴量データとを結合して、特徴量データを生成し、生成した特徴量データを特徴量データストレージ128に格納する。
- [0246] ステップS 43において、ループ判定部129は、特徴量データストレージ

ジ 1 2 8 に格納された特徴量データに対応する、セッション内特徴量データと、セッション間特徴量データとのそれぞれの特徴量毎の有効度スコアに基づいて、特徴量データの全体の有効度スコアを算出し、所定値以上か、または、処理開始からの経過時間が所定時間を経過したか否かを判定する。

[0247] ステップ S 4 3 において、特徴量データの全体の有効度スコアが所定値よりも小さく、かつ、処理開始からの経過時間が所定時間を経過していないと判定された場合、処理は、ステップ S 4 4 に進む。

[0248] ステップ S 4 4 において、ループ判定部 1 2 9 は、生成元選択処理において利用される精度閾値、および、有効度スコアに対して設定されるスコア閾値を所定値から小さくさせるように生成元選択部 1 2 3、および特徴量選択部 1 2 5 を制御して、処理は、ステップ S 3 5 に戻り、再度、特徴量データの生成処理を実行させる。

[0249] すなわち、ステップ S 4 3 において、特徴量データの全体の有効度スコアが所定値よりも小さく、かつ、処理開始からの経過時間が所定時間を経過していない場合、除外した系列データや特徴量にも有効なものが存在する可能性があるため、精度閾値およびスコア閾値の設定を所定値だけ小さくさせて、再度、特徴量データを生成させる。

[0250] ただし、この場合、この処理までに生成された特徴量データについては、特徴量データストレージ 1 2 8 に格納されたままの状態とし、以降においても有効であるものとする。また、以降においては、既に、特徴量データとして生成された特徴量は生成済みとして扱い、これまでの処理で、除外された生成元となる系列データや特徴量を復活させるようにして、再度、特徴量データが生成されるようにする。例えば、特徴量選択部 1 2 5 における有効度スコア計算において、ストレージ 1 2 8 に格納された特徴量、及び新規生成特徴量との和集合を用いてそれぞれ機械学習モデルを作成し、その精度の改善幅を新たな有効度スコアとして計算してもよい。

[0251] そして、ステップ S 4 3 において、特徴量データの全体の有効度スコアが所定値以上か、または、処理開始からの経過時間が所定時間を経過したと判

定された場合、処理は、ステップS 4 5に進む。

[0252] ステップS 4 5において、ループ判定部1 2 9は、特徴量データストレージ1 2 8に格納されている特徴量データのうち、特徴量データの全体の有効度スコアが最も高い特徴量データを読み出して、UI制御部6 1に出力してユーザに提示させると共に、機械学習モデル生成部6 3に出力する。

[0253] これに応じて、UI制御部6 1の生成特徴量可視化部1 0 3は、フローデータと特徴量データとに基づいて、UIを生成して、ユーザに提示する。

[0254] 尚、最初の処理で、ステップS 4 3の処理により、特徴量データの全体の有効度スコアが所定値よりも小さいまま、処理開始からの経過時間が所定時間を経過したと判定された場合、特徴量データの有効度スコアが不十分であり、特徴量データに基づいて生成される機械学習モデルによる予測精度が不十分である可能性があるので、生成特徴量可視化部1 0 3は、現状の有効度スコアを提示する共に、現状の特徴量データでは、予測精度が不十分な可能性があることを提示するようにしてもよい。

[0255] <生成元選択処理>

次に、図2 0のフローチャートを参照して、生成元選択部1 2 3による生成元選択処理について説明する。

[0256] ステップS 7 1において、生成元選択部1 2 3は、フローデータより出力フォーマットに基づいて抽出される系列データのうち、時間の経過に伴って変化のない時系列データなど、予測対象の予測に対して無関係とみられる系列データを除外する。

[0257] ステップS 7 2において、生成元選択部1 2 3は、系列データ毎に部分系列を取得して、所定の統計量からなる特徴量テーブルを作成する。

[0258] ステップS 7 3において、生成元選択部1 2 3は、系列データ毎に特徴量テーブルに基づいて、予測対象を予測する予測モデルを生成する。

[0259] ステップS 7 4において、生成元選択部1 2 3は、系列データ毎に予測モデルに基づいた予測結果の予測精度を算出する。

[0260] ステップS 7 5において、生成元選択部1 2 3は、予測モデルに基づいた

予測結果の予測精度が、所定の精度閾値よりも高い系列データを、セッション内特徴量の生成元として選択し、セッション内特徴量生成部124に出力する。

[0261] すなわち、以上の処理により、フローデータより出力フォーマットに基づいて抽出される系列データのうち、予測対象を予測するのに有効性の高い系列データをセッション内特徴量の生成元の系列データとして選択して、セッション内特徴量生成部124に出力することが可能となる。

[0262] 結果として、セッション内特徴量データおよびセッション間特徴量データからなる特徴量データに基づいて、機械学習により生成される機械学習モデルの予測精度を向上させることが可能となる。

[0263] <セッション内特徴量データ生成処理>

次に、図21のフローチャートを参照して、セッション内特徴量生成部124によるセッション内特徴量データ生成処理について説明する。

[0264] ステップS91において、セッション内特徴量生成部124は、メタデータ抽出部124aを制御して、フローデータよりメタデータを抽出して、生成させる。

[0265] ステップS92において、セッション内特徴量生成部124は、推定モデル124bを用いて、メタデータから有効特徴量の作成方法を推定させる。

[0266] ステップS93において、セッション内特徴量生成部124は、推定モデル124bにより推定された有効特徴量の作成方法に基づいて、生成元選択部123より供給されるセッション内特徴量の生成元として選択された系列データを利用して、特徴量を生成し、生成した特徴量に基づいて、セッション内特徴量データを生成し、特徴量選択部125に出力する。

[0267] 以上の処理により、セッション内特徴量データを構成する特徴量は、生成元選択部123において、フローデータより抽出された系列データのうち、予測対象の予測に有効とされる系列データを生成元として利用した上で、さらに、フローデータから生成されるメタデータに基づいて推定された有効特徴量の生成方法を用いて生成される。

- [0268] また、上述したように、セッション内特徴量データを構成する特徴量のうち、特徴量選択部125において、さらに、有効度スコアが求められ、有効度スコアが所定のスコア閾値よりも高いものだけが選択されて、セッション内特徴量データが生成される。
- [0269] さらに、このセッション内特徴量データに基づいて、セッション間特徴量データが生成され、このセッション間特徴量データを構成する特徴量においても、有効度スコアが所定のスコア閾値よりも高いものが選択されて、セッション間特徴量データが再構成される。
- [0270] すなわち、予測対象の予測に係る有効度スコアの高い特徴量からなるセッション内特徴量データと、そのセッション内特徴量データに基づいて、セッション間特徴量データが生成された上で、さらに、有効度スコアに基づいた特徴量を選択されて、セッション間特徴量データが生成される。
- [0271] そして、このようにして生成されたセッション内特徴量データと、セッション間特徴量データとが結合されて特徴量データが生成されるので、予測対象の予測に有効性の高い特徴量データを生成することが可能となる。
- [0272] また、生成された特徴量データの全体の有効度スコアが所定の閾値よりも高く、予測対象の予測に十分であると認められるときには、設定された処理時間内である限り、より多くの除外した系列データや特徴量にも有効なものが存在する可能性があるので、精度閾値およびスコア閾値の設定を所定値だけ小さくさせて、再度、特徴量データを生成させる。
- [0273] 結果として、予測対象を予測する機械学習モデルの生成に使用する、高い有効性を備えた特徴量データを、より多く生成することが可能となる。
- [0274] <変形例>
- セッションをクラスタリングすることにより、例えば、設定されたセッションの上位となる集合を作成するようにしてもよい。
- [0275] 例えば、セッションを設定する際に、予めセッションをクラスタリングした上で、セッションよりも上位集合を設定し、その上位となる集合ごとにセッションが設定されるようにしてもよい。

- [0276] 例えば、時系列データをシェープレット分解して、特徴的な部分波形の集合を離散化して、離散化された部分波形を単語とみなし、時系列データやセッションを文章とみなすようにして、TF-IDF (Term Frequency-Inverse Document Frequency) 値を求めて、セッションの上位となる集合を求めるようにしてもよい。
- [0277] すなわち、例えば、図22の上部で示されるようなセッションFW1乃至FW3が存在するような場合について考える。
- [0278] ここで、セッションFW1を特徴的な部分波形PW1-1, PW2-1, PW3-1からなる集合とみなし、セッションFW2を特徴的な部分波形PW1-11, PW3-11, PW3-12からなる集合とみなし、セッションFW3を特徴的な部分波形PW2-21, PW1-21からなる集合とみなし、それぞれを離散化し、部分波形についてTF-IDFを実行する。
- [0279] 図22の下部においては、セッションFW1の(PW1, PW2, PW3)のTF-IDF値が、(0, 0.1353, 0.1353)とされ、セッションFW2の(PW1, PW2, PW3)のTF-IDF値が、(0, 0, 0.2706)とされ、セッションFW3の(PW1, PW2, PW3)のTF-IDF値が、(0, 0.2050, 0)とされている。
- [0280] そして、セッション毎のTF-IDF値に基づくベクトルに基づいたクラスタリングにより、類似度が高いセッション同士を同一クラスにして、上位集合が設定されるようにしてもよい。
- [0281] また、図23の左部で示されるような、打席IDがセッションとして設定され、1セッション1行で設定されるような場合については、フローデータよりメタデータを抽出し、抽出したメタデータに基づいて、セッションである打席ID毎に、属性データの統計量、例えば、投手IDの頻度などから、セッションである打席IDのクラスタリングを実行することにより、セッションをグルーピングし、セッション上位集合列(図中クラスタID列)を新規作成するようにしてもよい。
- [0282] 図23の右部においては、例えば、フローデータのメタデータとして抽出

された、例えば、投手IDにより、セッションである打席IDで分類される打席毎の相手投手によりクラスタリングすることでクラスタIDとして、図中上からA, B, Aと分類される例が示されている。すなわち、ここでは、クラスタIDは、投手IDに対応するものとなる。

[0283] <<3. ソフトウェアにより実行させる例>>

ところで、上述した一連の処理は、ハードウェアにより実行させることもできるが、ソフトウェアにより実行させることもできる。一連の処理をソフトウェアにより実行させる場合には、そのソフトウェアを構成するプログラムが、専用のハードウェアに組み込まれているコンピュータ、または、各種のプログラムをインストールすることで、各種の機能を実行することが可能な、例えば汎用のコンピュータなどに、記録媒体からインストールされる。

[0284] 図24は、汎用のコンピュータの構成例を示している。このコンピュータは、CPU(Central Processing Unit)1001を内蔵している。CPU1001にはバス1004を介して、入出力インタフェース1005が接続されている。バス1004には、ROM(Read Only Memory)1002およびRAM(Random Access Memory)1003が接続されている。

[0285] 入出力インタフェース1005には、ユーザが操作コマンドを入力するキーボード、マウスなどの入力デバイスよりなる入力部1006、処理操作画面や処理結果の画像を表示デバイスに出力する出力部1007、プログラムや各種データを格納するハードディスクドライブなどよりなる記憶部1008、LAN(Local Area Network)アダプタなどよりなり、インターネットに代表されるネットワークを介した通信処理を実行する通信部1009が接続されている。また、磁気ディスク(フレキシブルディスクを含む)、光ディスク(CD-ROM(Compact Disc-Read Only Memory)、DVD(Digital Versatile Disc)を含む)、光磁気ディスク(MD(Mini Disc)を含む)、もしくは半導体メモリなどのリムーバブル記憶媒体1011に対してデータを読み書きするドライブ1010が接続されている。

[0286] CPU1001は、ROM1002に記憶されているプログラム、または磁気デ

ィスク、光ディスク、光磁気ディスク、もしくは半導体メモリ等のリムーバブル記憶媒体1011ら読み出されて記憶部1008にインストールされ、記憶部1008からRAM1003にロードされたプログラムに従って各種の処理を実行する。RAM1003にはまた、CPU1001が各種の処理を実行する上において必要なデータなども適宜記憶される。

[0287] 以上のように構成されるコンピュータでは、CPU1001が、例えば、記憶部1008に記憶されているプログラムを、入出力インタフェース1005及びバス1004を介して、RAM1003にロードして実行することにより、上述した一連の処理が行われる。

[0288] コンピュータ（CPU1001）が実行するプログラムは、例えば、パッケージメディア等としてのリムーバブル記憶媒体1011に記録して提供することができる。また、プログラムは、ローカルエリアネットワーク、インターネット、デジタル衛星放送といった、有線または無線の伝送媒体を介して提供することができる。

[0289] コンピュータでは、プログラムは、リムーバブル記憶媒体1011をドライブ1010に装着することにより、入出力インタフェース1005を介して、記憶部1008にインストールすることができる。また、プログラムは、有線または無線の伝送媒体を介して、通信部1009で受信し、記憶部1008にインストールすることができる。その他、プログラムは、ROM1002や記憶部1008に、あらかじめインストールしておくことができる。

[0290] なお、コンピュータが実行するプログラムは、本明細書で説明する順序に沿って時系列に処理が行われるプログラムであっても良いし、並列に、あるいは呼び出しが行われたとき等の必要なタイミングで処理が行われるプログラムであっても良い。

[0291] 尚、図24におけるCPU1001が、図2の情報処理装置31の制御部51の機能を実現させる。

[0292] また、本明細書において、システムとは、複数の構成要素（装置、モジュール（部品）等）の集合を意味し、すべての構成要素が同一筐体中にあるか

否かは問わない。したがって、別個の筐体に収納され、ネットワークを介して接続されている複数の装置、及び、1つの筐体の中に複数のモジュールが収納されている1つの装置は、いずれも、システムである。

[0293] なお、本開示の実施の形態は、上述した実施の形態に限定されるものではなく、本開示の要旨を逸脱しない範囲において種々の変更が可能である。

[0294] 例えば、本開示は、1つの機能をネットワークを介して複数の装置で分担、共同して処理するクラウドコンピューティングの構成をとることができる。

[0295] また、上述のフローチャートで説明した各ステップは、1つの装置で実行する他、複数の装置で分担して実行することができる。

[0296] さらに、1つのステップに複数の処理が含まれる場合には、その1つのステップに含まれる複数の処理は、1つの装置で実行する他、複数の装置で分担して実行することができる。

[0297] 尚、本開示は、以下のような構成も取ることができる。

<1> 少なくとも時系列データを含むフローデータのメタデータを生成するメタデータ生成部と、

前記メタデータに基づいて、前記フローデータを構成する系列データより特徴量の生成方法を推定する推定部と、

前記推定部により推定された生成方法で、前記系列データより特徴量を生成する特徴量生成部と

を備える情報処理装置。

<2> 前記フローデータにおける、セッション単位、時刻単位、および予測対象の設定を受け付ける設定部をさらに含み、

前記メタデータ生成部は、前記フローデータのうち、前記セッション単位、前記時刻単位、および前記予測対象の設定に応じて、前記フローデータより抽出された前記系列データより前記メタデータを生成する

<1>に記載の情報処理装置。

<3> 前記フローデータを構成する列を推定する列推定部をさらに含み、

前記設定部は、前記列推定部により推定された列を提示して、前記列を単位として、前記フローデータにおける、前記セッション単位、前記時刻単位、および前記予測対象の列の設定を促すUI (User Interface) 画像を生成して提示し、前記UI画像に基づいて、前記セッション単位、前記時刻単位、および前記予測対象の列の設定を受け付ける

<2>に記載の情報処理装置。

<4> 前記設定部により設定された、前記フローデータにおける、前記セッション単位、前記時刻単位、および前記予測対象に基づいて、前記フローデータより抽出する系列データの出力フォーマットを決定する出力フォーマット決定部をさらに含み、

前記メタデータ生成部は、前記フローデータのうち、前記セッション単位、前記時刻単位、および前記予測対象の設定に応じて決定された前記出力フォーマットに基づいて、前記フローデータより抽出された前記系列データより前記メタデータを生成する

<2>に記載の情報処理装置。

<5> 前記出力フォーマットに基づいて、前記フローデータより抽出された前記系列データのそれぞれについての、前記予測対象の予測に係る予測精度を求め、所定の精度閾値よりも高い系列データを選択する選択部をさらに備え、

前記メタデータ生成部は、前記出力フォーマットに基づいて、前記フローデータより抽出された前記系列データのうち、前記選択部により選択された前記系列データより前記メタデータを生成する

<4>に記載の情報処理装置。

<6> 前記選択部は、前記出力フォーマットに基づいて、前記フローデータより抽出された系列データのそれぞれについて、部分系列毎の特徴量を求め、前記予測対象を予測するための予測モデルに、前記部分系列毎の特徴量を入力することで、前記予測対象を予測し、前記予測対象と、前記予測モデルによる予測結果との比較から、前記系列データ毎の、前記予測対象の予測

に係る予測精度を求め、前記所定の精度閾値よりも高い系列データを選択する

<5>に記載の情報処理装置。

<7> 前記特徴量生成部は、前記推定部により推定された前記特徴量の生成方法で、前記系列データより特徴量を生成し、生成した前記セッション単位の特徴量に基づいて、セッション内特徴量を生成する

<2>に記載の情報処理装置。

<8> 前記セッション内特徴量を構成する特徴量のそれぞれについて、前記予測対象の予測に対する有効度スコアを算出する有効度スコア算出部と、

前記セッション内特徴量を構成する特徴量のうち、前記有効度スコアに基づいて、所定のスコア閾値より高い特徴量を選択して、前記セッション内特徴量を再構成するセッション内特徴量選択部をさらに含む

<7>に記載の情報処理装置。

<9> 前記セッション内特徴量に基づいて、前記セッション間の特徴量を含む、セッション間特徴量を生成するセッション間特徴量生成部をさらに含む

<8>に記載の情報処理装置。

<10> 前記有効度スコアは、前記セッション間特徴量を構成する特徴量のそれぞれについても、前記予測対象の予測に対する有効度スコアを算出し、

前記セッション間特徴量を構成する特徴量のうち、前記有効度スコアに基づいて、所定のスコア閾値より高い特徴量を選択して、前記セッション間特徴量を再構成するセッション間特徴量選択部をさらに含む

<9>に記載の情報処理装置。

<11> 前記有効度スコア算出部は、前記セッション内特徴量、および前記セッション間特徴量を構成する特徴量のそれぞれと、前記予測対象との相互情報量を前記有効度スコアとして算出する

<10>に記載の情報処理装置。

< 1 2 > 前記有効度スコア算出部は、前記セッション内特徴量、および前記セッション間特徴量を構成する特徴量に基づいて簡易的に生成される機械学習モデルにより、前記予測対象を予測する予測精度を前記有効度スコアとして算出し、

前記セッション内特徴量選択部は、前記有効度スコアが所定のスコア閾値よりも低くならない、前記特徴量の部分集合を選択して、前記セッション内特徴量を再構成し、

前記セッション間特徴量選択部は、前記有効度スコアが所定のスコア閾値よりも低くならない、前記特徴量の部分集合を選択して、前記セッション間特徴量を再構成する

< 1 0 >に記載の情報処理装置。

< 1 3 > 再構成された前記セッション内特徴量、および再構成された前記セッション間特徴量を結合する結合部と、

前記結合部により結合された、再構成された前記セッション内特徴量、および再構成された前記セッション間特徴量のそれぞれの特徴量の有効度スコアに基づいて、前記結合部により結合された特徴量の全体の有効度スコアを算出し、前記全体の有効度スコアが所定の閾値よりも小さいか否かを判定する判定部をさらに含み、

前記判定部は、前記全体の有効度スコアが所定の閾値よりも小さいとき、前記スコア閾値を所定値だけ小さくして、前記セッション内特徴量選択部、および前記セッション間特徴量選択部による処理を再度実行させる

< 1 0 >に記載の情報処理装置。

< 1 4 > 前記推定部は、前記フローデータの前記メタデータと、前記フローデータより抽出された系列データより生成される、所定の機械学習モデルの学習に用いた特徴量の作成方法の分布とをペアの情報とし、前記ペアの情報に基づいた学習により生成された推定モデルであり、前記メタデータに基づいて、前記特徴量の生成方法を推定する

< 1 >乃至< 1 3 >のいずれかに記載の情報処理装置。

<15> 前記フローデータは、時間の経過に対して変化する、前記時系列データに加えて、前記時間の経過に対して不変なデータからなる属性データをさらに含む

<1>乃至<14>のいずれかに記載の情報処理装置。

<16> 少なくとも時系列データを含むフローデータのメタデータを生成し、

前記メタデータに基づいて、前記フローデータを構成する系列データより特徴量の生成方法を推定し、

推定された生成方法で、前記系列データより特徴量を生成するステップを含む情報処理方法。

<17> 少なくとも時系列データを含むフローデータのメタデータを生成するメタデータ生成部と、

前記メタデータに基づいて、前記フローデータを構成する系列データより特徴量の生成方法を推定する推定部と、

前記推定部により推定された生成方法で、前記系列データより特徴量を生成する特徴量生成部と

してコンピュータを機能させるプログラム。

符号の説明

[0298] 31 情報処理装置, 61 UI制御部, 62 データ処理部, 63 機械学習モデル生成部, 101 フローデータ入力部, 102 タスク設定部, 103 生成特徴量可視化部, 121 列推定部, 122 出力フォーマット決定部, 123 生成元選択部, 124 セッション内特徴量生成部, 124 a メタデータ抽出部, 124 b 推定モデル, 125 特徴量選択部, 126 セッション間特徴量生成部, 127 結合部, 128 特徴量データストレージ, 129 ループ判定部

請求の範囲

- [請求項1] 少なくとも時系列データを含むフローデータのメタデータを生成するメタデータ生成部と、
- 前記メタデータに基づいて、前記フローデータを構成する系列データより特徴量の生成方法を推定する推定部と、
- 前記推定部により推定された生成方法で、前記系列データより特徴量を生成する特徴量生成部と
- を備える情報処理装置。
- [請求項2] 前記フローデータにおける、セッション単位、時刻単位、および予測対象の設定を受け付ける設定部をさらに含み、
- 前記メタデータ生成部は、前記フローデータのうち、前記セッション単位、前記時刻単位、および前記予測対象の設定に応じて、前記フローデータより抽出された前記系列データより前記メタデータを生成する
- 請求項1に記載の情報処理装置。
- [請求項3] 前記フローデータを構成する列を推定する列推定部をさらに含み、
- 前記設定部は、前記列推定部により推定された列を提示して、前記列を単位として、前記フローデータにおける、前記セッション単位、前記時刻単位、および前記予測対象の列の設定を促すUI (User Interface) 画像を生成して提示し、前記UI画像に基づいて、前記セッション単位、前記時刻単位、および前記予測対象の列の設定を受け付ける
- 請求項2に記載の情報処理装置。
- [請求項4] 前記設定部により設定された、前記フローデータにおける、前記セッション単位、前記時刻単位、および前記予測対象に基づいて、前記フローデータより抽出する系列データの出力フォーマットを決定する出力フォーマット決定部をさらに含み、
- 前記メタデータ生成部は、前記フローデータのうち、前記セッション単位、前記時刻単位、および前記予測対象の設定に応じて決定され

た前記出力フォーマットに基づいて、前記フローデータより抽出された前記系列データより前記メタデータを生成する

請求項2に記載の情報処理装置。

[請求項5] 前記出力フォーマットに基づいて、前記フローデータより抽出された前記系列データのそれぞれについての、前記予測対象の予測に係る予測精度を求め、所定の精度閾値よりも高い系列データを選択する選択部をさらに備え、

前記メタデータ生成部は、前記出力フォーマットに基づいて、前記フローデータより抽出された前記系列データのうち、前記選択部により選択された前記系列データより前記メタデータを生成する

請求項4に記載の情報処理装置。

[請求項6] 前記選択部は、前記出力フォーマットに基づいて、前記フローデータより抽出された系列データのそれぞれについて、部分系列毎の特徴量を求め、前記予測対象を予測するための予測モデルに、前記部分系列毎の特徴量を入力することで、前記予測対象を予測し、前記予測対象と、前記予測モデルによる予測結果との比較から、前記系列データ毎の、前記予測対象の予測に係る予測精度を求め、前記所定の精度閾値よりも高い系列データを選択する

請求項5に記載の情報処理装置。

[請求項7] 前記特徴量生成部は、前記推定部により推定された前記特徴量の生成方法で、前記系列データより特徴量を生成し、生成した前記セッション単位の特徴量に基づいて、セッション内特徴量を生成する

請求項2に記載の情報処理装置。

[請求項8] 前記セッション内特徴量を構成する特徴量のそれぞれについて、前記予測対象の予測に対する有効度スコアを算出する有効度スコア算出部と、

前記セッション内特徴量を構成する特徴量のうち、前記有効度スコアに基づいて、所定のスコア閾値より高い特徴量を選択して、前記セ

セッション内特徴量を再構成するセッション内特徴量選択部をさらに含む

請求項7に記載の情報処理装置。

[請求項9] 前記セッション内特徴量に基づいて、前記セッション間の特徴量を含む、セッション間特徴量を生成するセッション間特徴量生成部をさらに含む

請求項8に記載の情報処理装置。

[請求項10] 前記有効度スコアは、前記セッション間特徴量を構成する特徴量のそれぞれについても、前記予測対象の予測に対する有効度スコアを算出し、

前記セッション間特徴量を構成する特徴量のうち、前記有効度スコアに基づいて、所定のスコア閾値より高い特徴量を選択して、前記セッション間特徴量を再構成するセッション間特徴量選択部をさらに含む

請求項9に記載の情報処理装置。

[請求項11] 前記有効度スコア算出部は、前記セッション内特徴量、および前記セッション間特徴量を構成する特徴量のそれぞれと、前記予測対象との相互情報量を前記有効度スコアとして算出する

請求項10に記載の情報処理装置。

[請求項12] 前記有効度スコア算出部は、前記セッション内特徴量、および前記セッション間特徴量を構成する特徴量に基づいて簡易的に生成される機械学習モデルにより、前記予測対象を予測する予測精度を前記有効度スコアとして算出し、

前記セッション内特徴量選択部は、前記有効度スコアが所定のスコア閾値よりも低くならない、前記特徴量の部分集合を選択して、前記セッション内特徴量を再構成し、

前記セッション間特徴量選択部は、前記有効度スコアが所定のスコア閾値よりも低くならない、前記特徴量の部分集合を選択して、前記

セッション間特徴量を再構成する

請求項10に記載の情報処理装置。

[請求項13] 再構成された前記セッション内特徴量、および再構成された前記セッション間特徴量を結合する結合部と、

前記結合部により結合された、再構成された前記セッション内特徴量、および再構成された前記セッション間特徴量のそれぞれの特徴量の有効度スコアに基づいて、前記結合部により結合された特徴量の全体の有効度スコアを算出し、前記全体の有効度スコアが所定の閾値よりも小さいか否かを判定する判定部をさらに含み、

前記判定部は、前記全体の有効度スコアが所定の閾値よりも小さいとき、前記スコア閾値を所定値だけ小さくして、前記セッション内特徴量選択部、および前記セッション間特徴量選択部による処理を再度実行させる

請求項10に記載の情報処理装置。

[請求項14] 前記推定部は、前記フローデータの前記メタデータと、前記フローデータより抽出された系列データより生成される、所定の機械学習モデルの学習に用いた特徴量の作成方法の分布とをペアの情報とし、前記ペアの情報に基づいた学習により生成された推定モデルであり、前記メタデータに基づいて、前記特徴量の生成方法を推定する

請求項1に記載の情報処理装置。

[請求項15] 前記フローデータは、時間の経過に対して変化する、前記時系列データに加えて、前記時間の経過に対して不変なデータからなる属性データをさらに含む

請求項1に記載の情報処理装置。

[請求項16] 少なくとも時系列データを含むフローデータのメタデータを生成し、
前記メタデータに基づいて、前記フローデータを構成する系列データより特徴量の生成方法を推定し、

推定された生成方法で、前記系列データより特徴量を生成するステップを含む情報処理方法。

[請求項17]

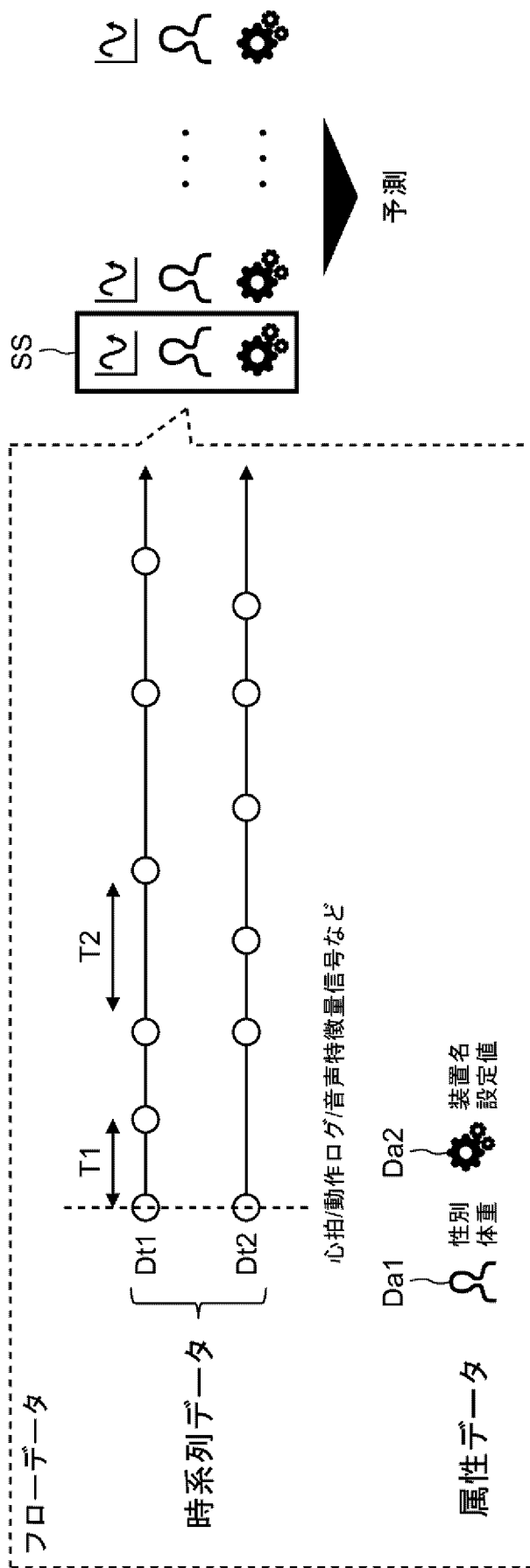
少なくとも時系列データを含むフローデータのメタデータを生成するメタデータ生成部と、

前記メタデータに基づいて、前記フローデータを構成する系列データより特徴量の生成方法を推定する推定部と、

前記推定部により推定された生成方法で、前記系列データより特徴量を生成する特徴量生成部と

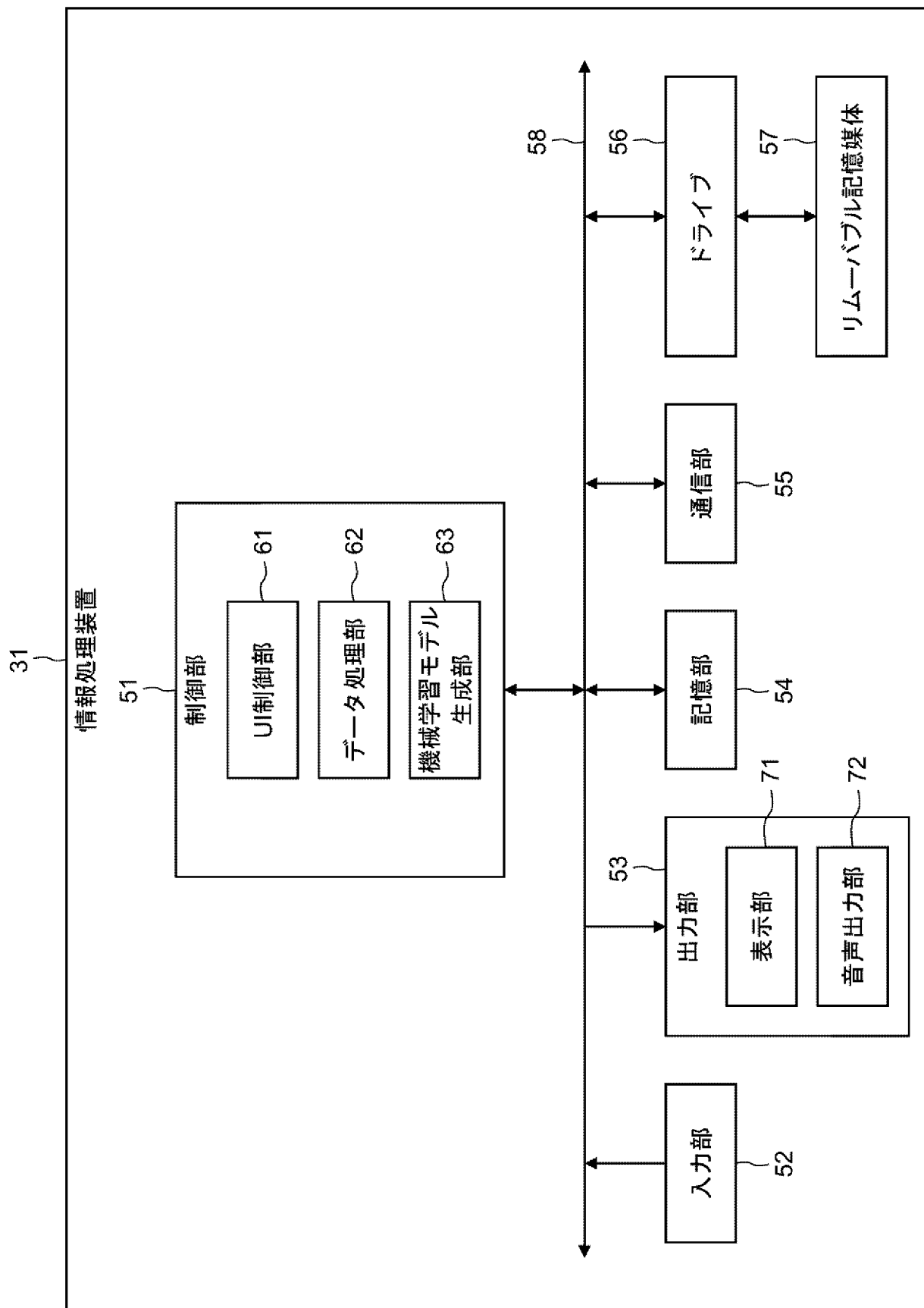
してコンピュータを機能させるプログラム。

[図1]
FIG.1

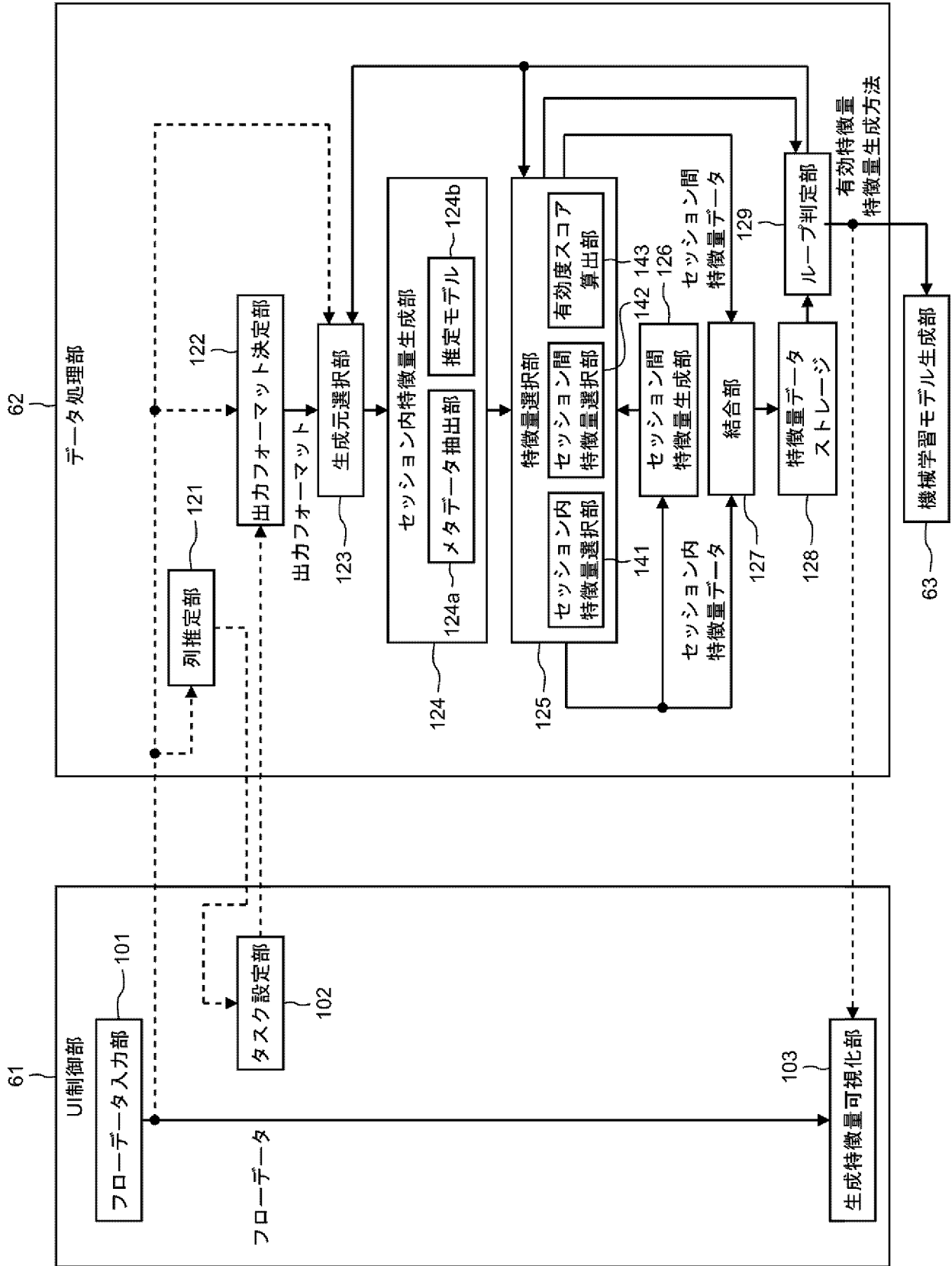


[図2]
FIG.2

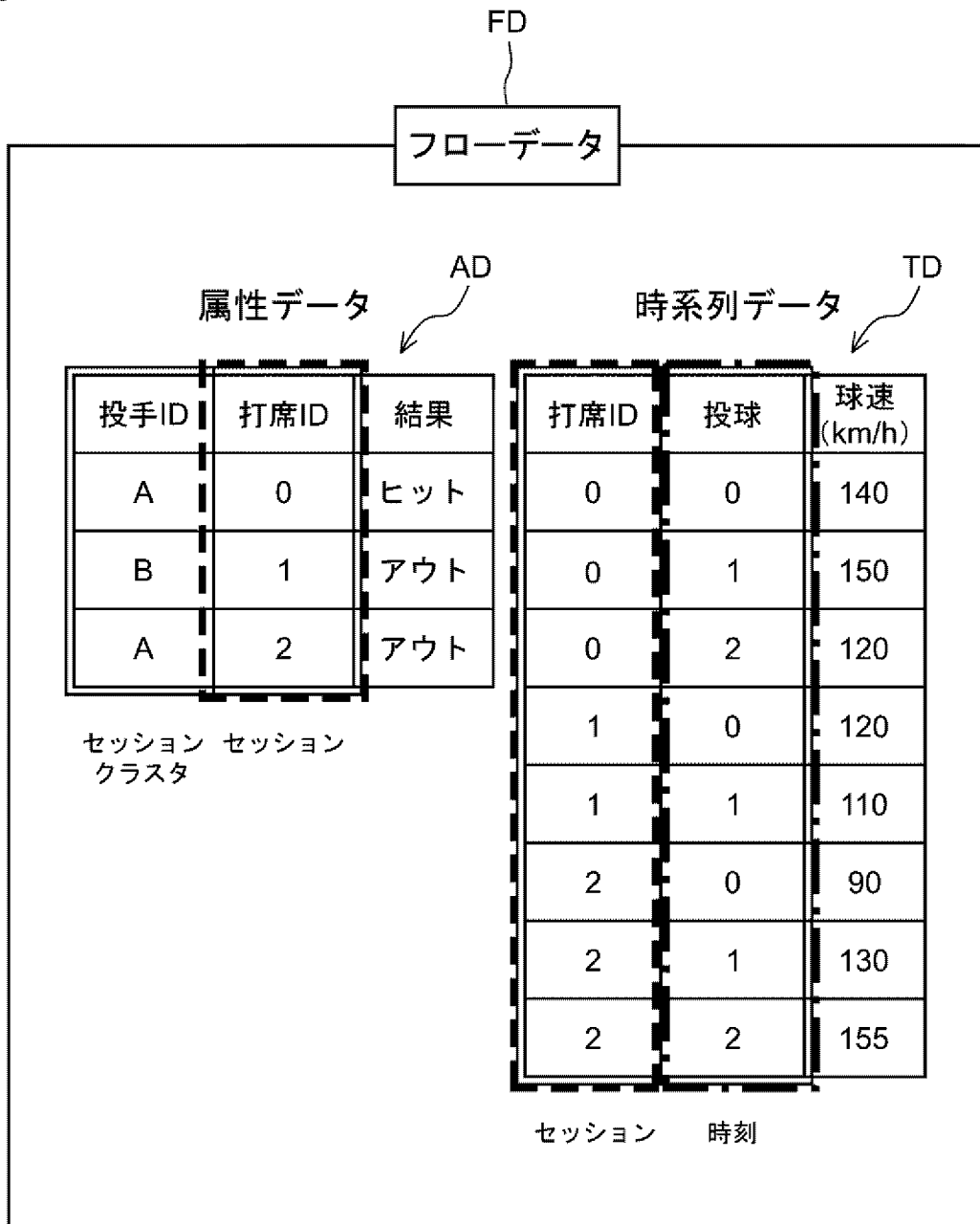
フローデータ	セッション単位	時刻単位	属性データ	時系列データ
病院バイタルログ	患者	日時	患者の性別	心拍番号
工場ロボットログ	ロボット	日時	ロボットの故障回数	トルクセンサ番号
野球投球ログ	打席	打席内球数	投手の左/右投げ	球速

[図3]
FIG.3

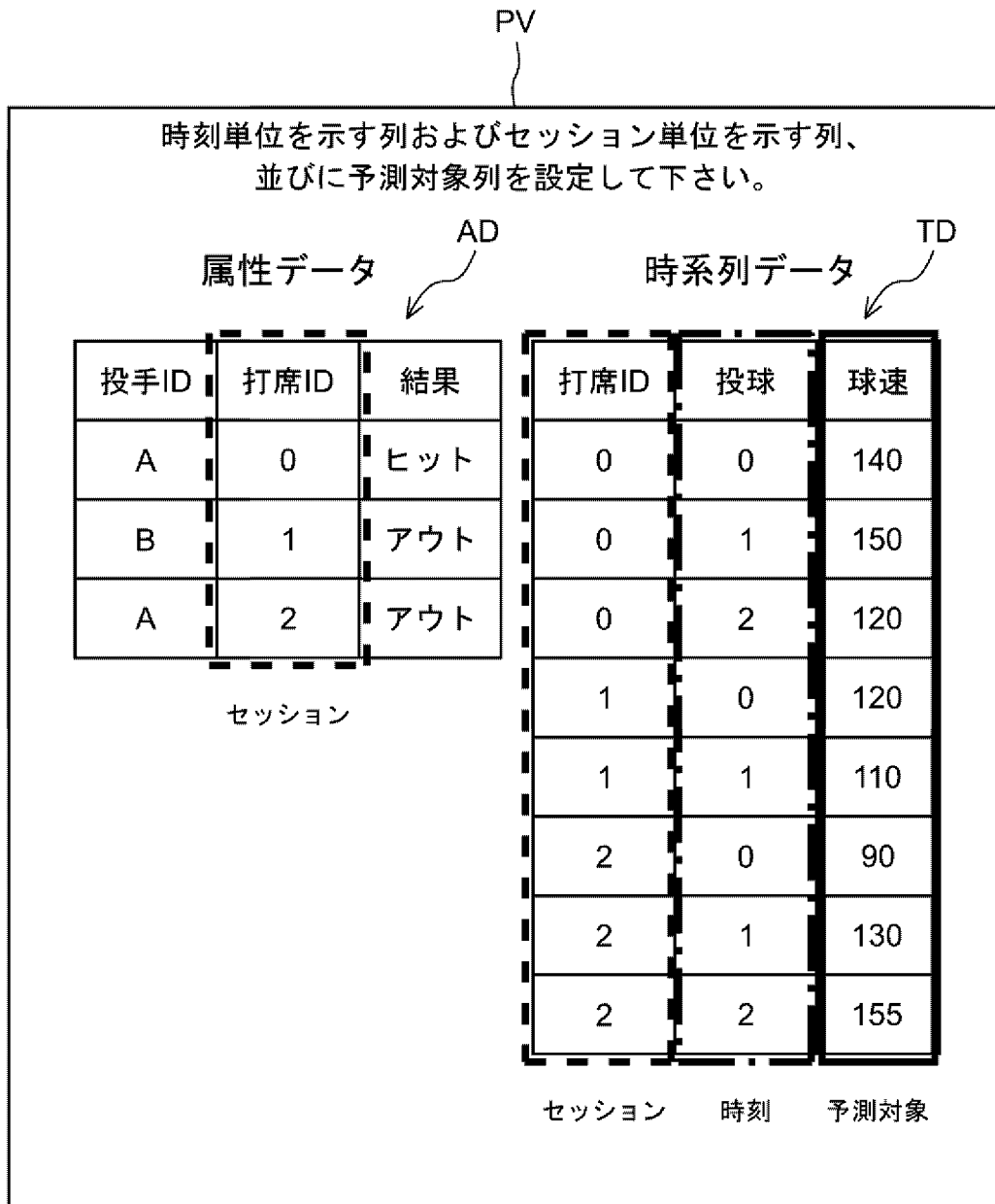
[図4]
FIG.4



[図5]
FIG.5



[図6]
FIG.6



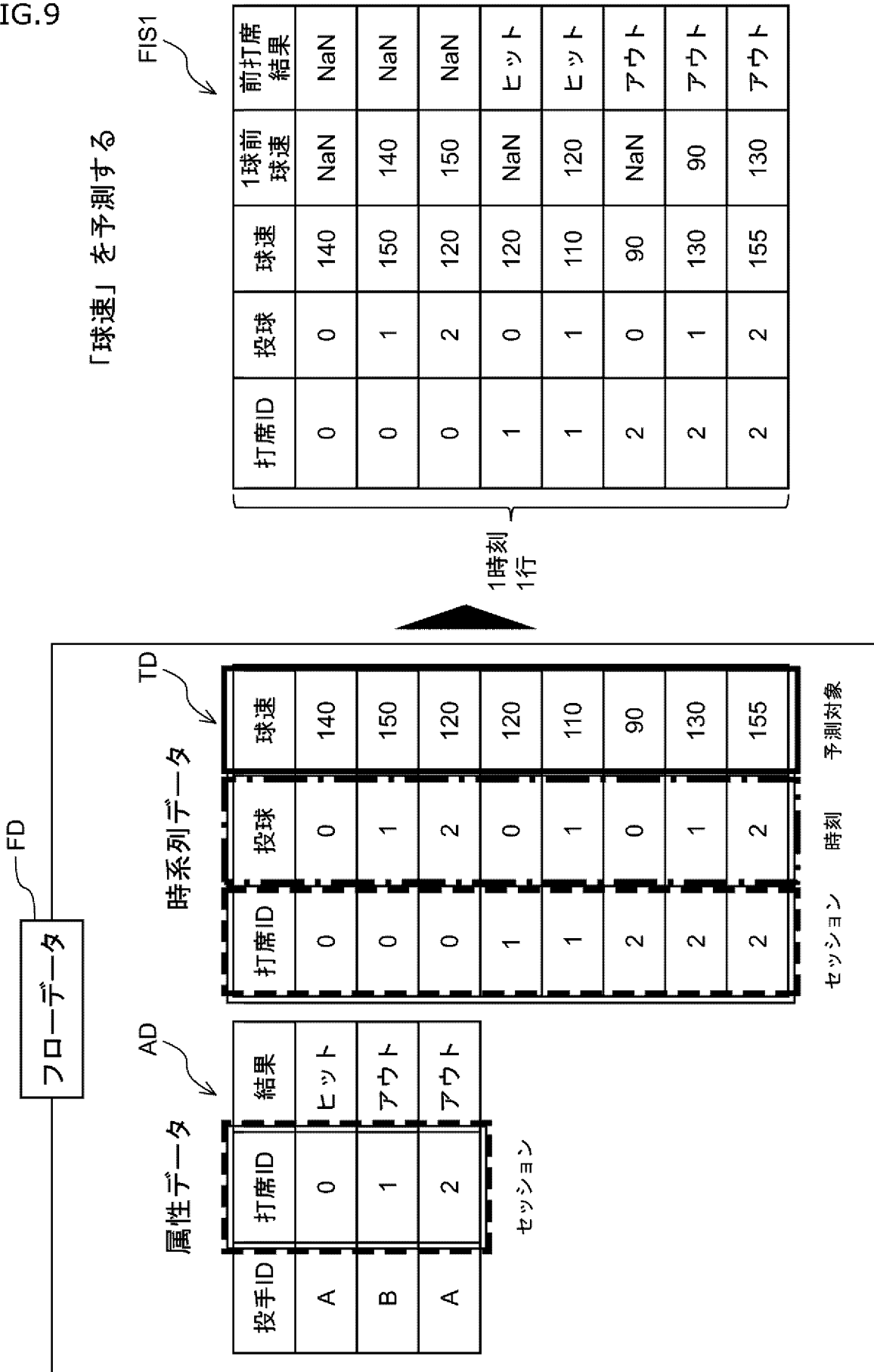
[図7]
FIG.7

id	time	name	value
A	t1	x	x(A,t1)
A	t2	x	x(A,t2)
A	t1	y	y(A,t1)
A	t2	y	y(A,t2)
B	t1	x	x(B,t1)
B	t2	x	x(B,t2)
B	t1	y	y(B,t1)
B	t2	y	y(B,t2)

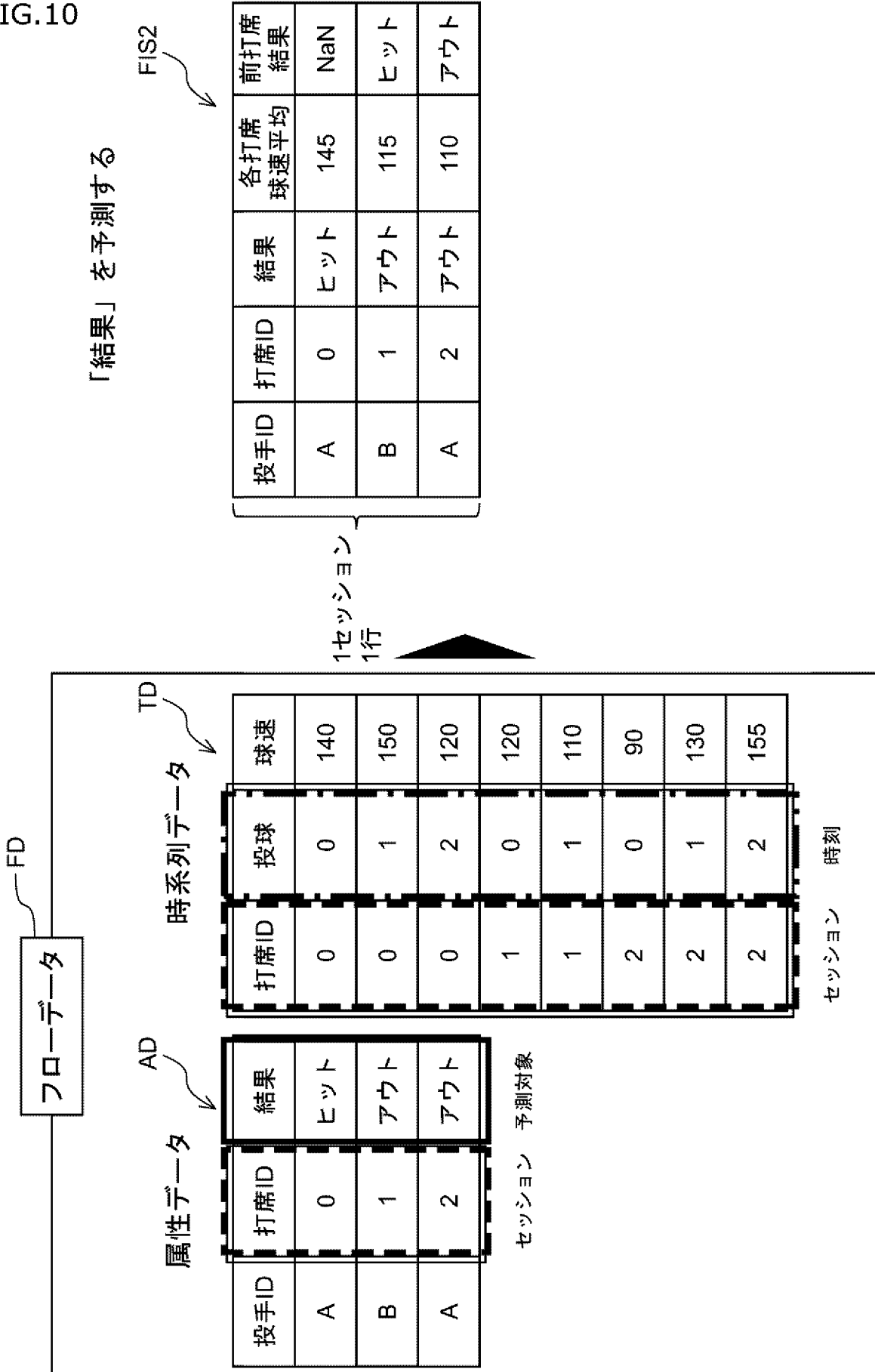
[図8]
FIG.8

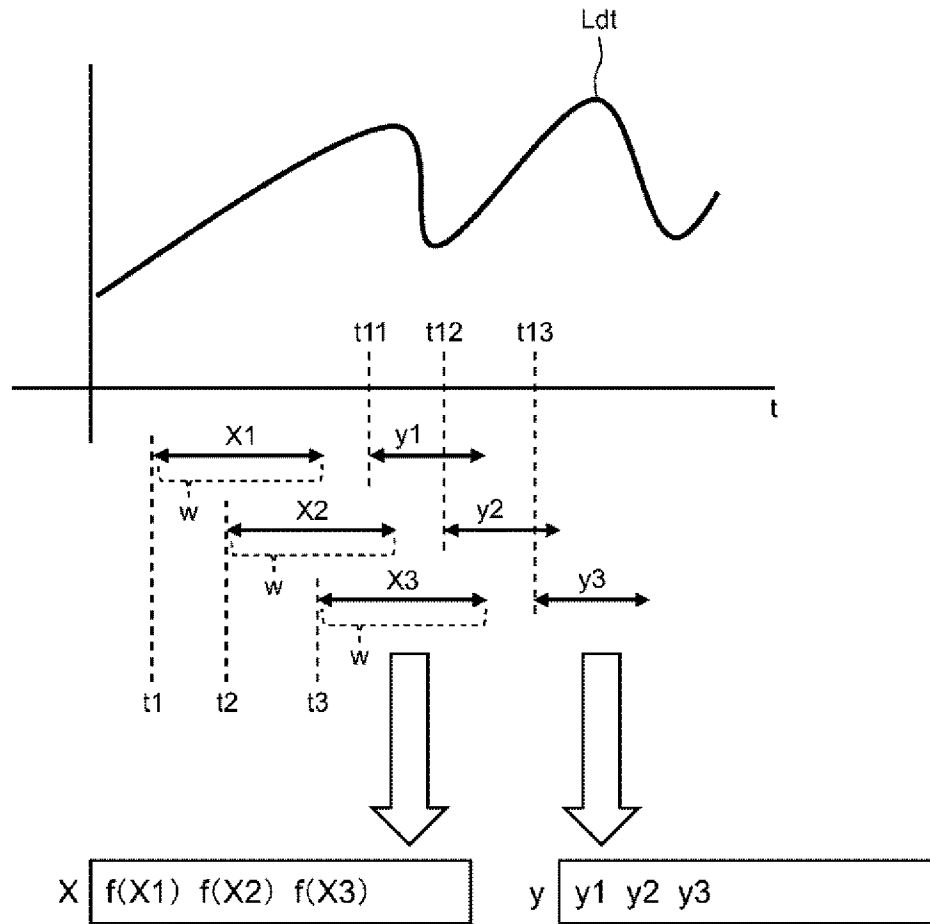
id	time	value x	value y
A	t1	x(A,t1)	y(A,t1)
A	t2	x(A,t2)	y(A,t2)
B	t1	x(B,t1)	y(B,t1)
B	t2	x(B,t2)	y(B,t2)

[図9]
FIG.9

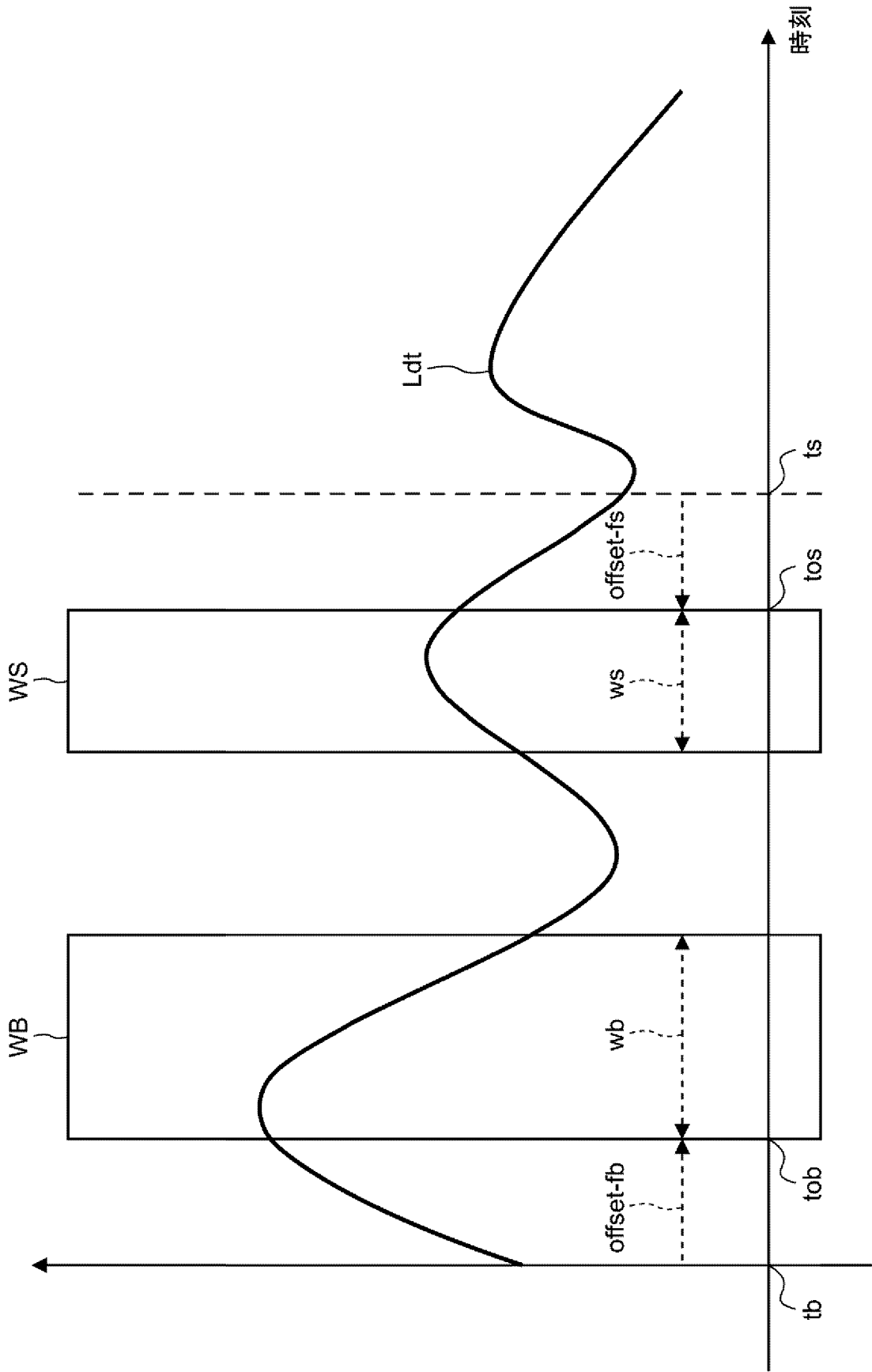


[図10]
FIG.10

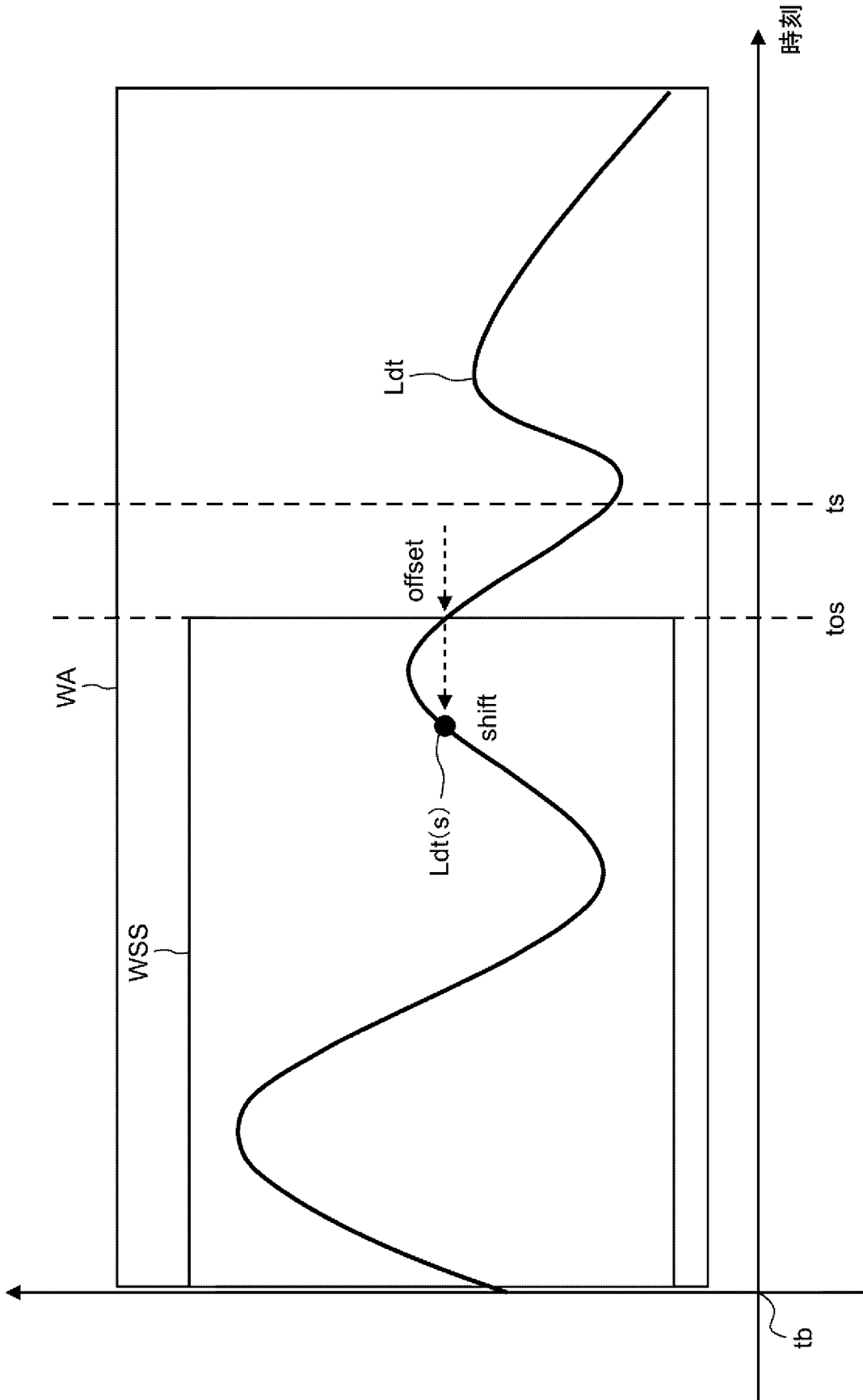


[図11]
FIG.11

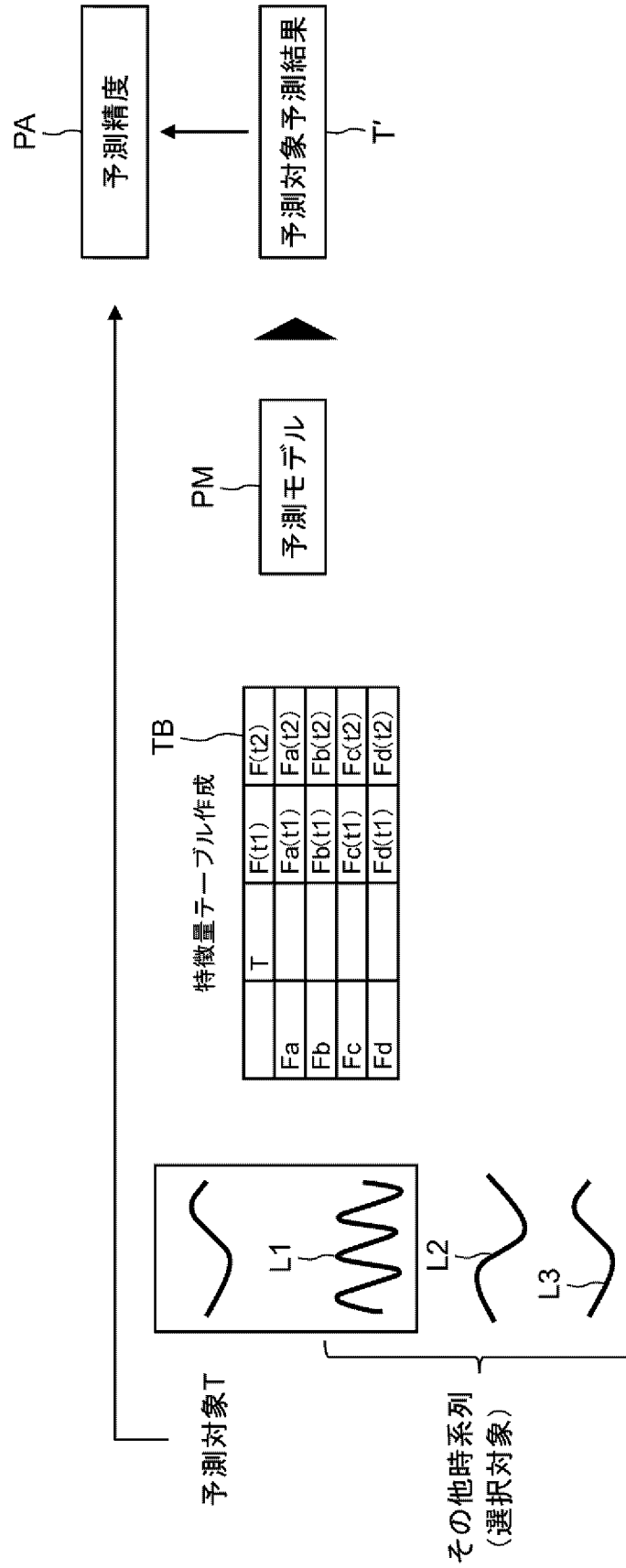
[図12]
FIG.12

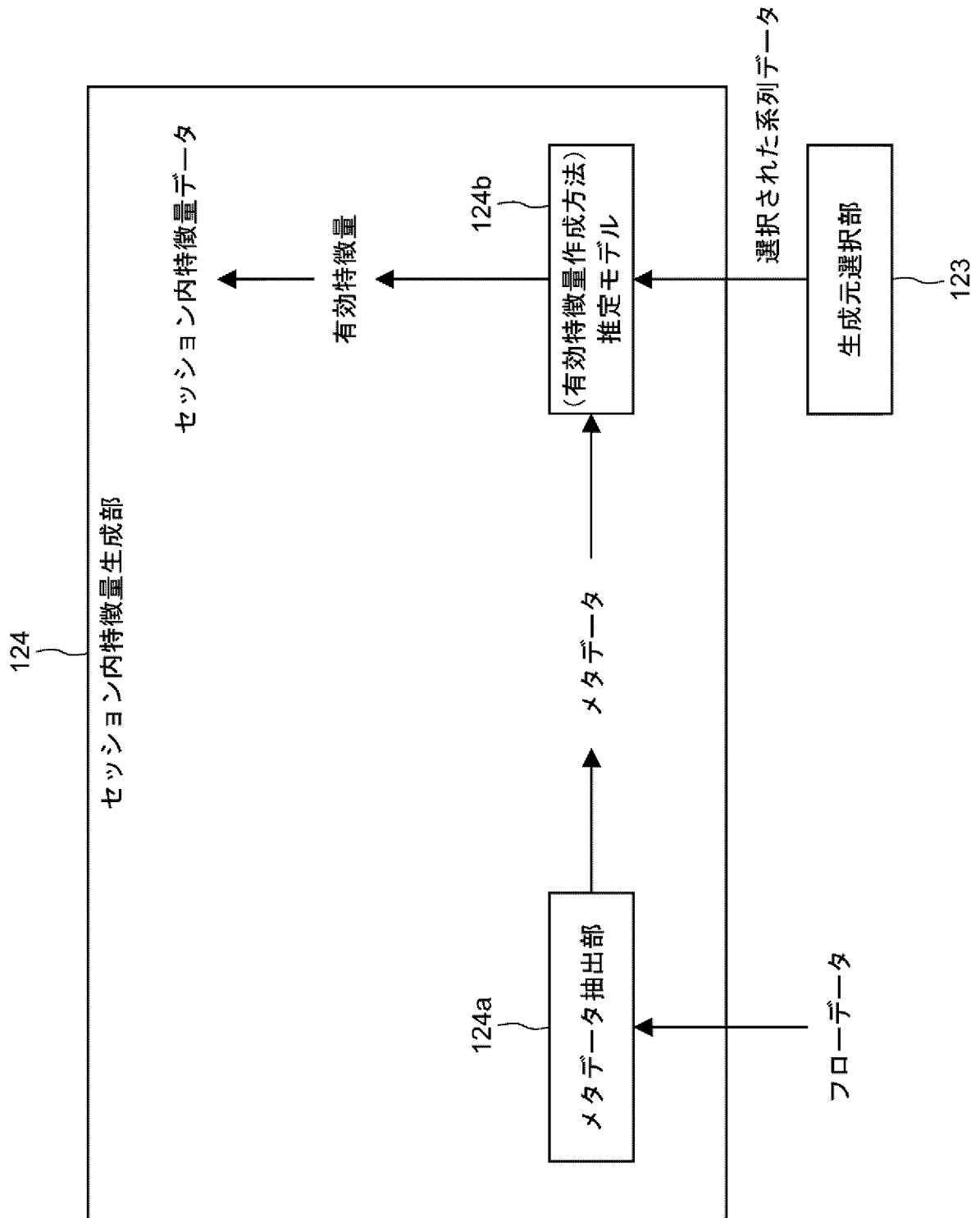


[圖13]
FIG.13

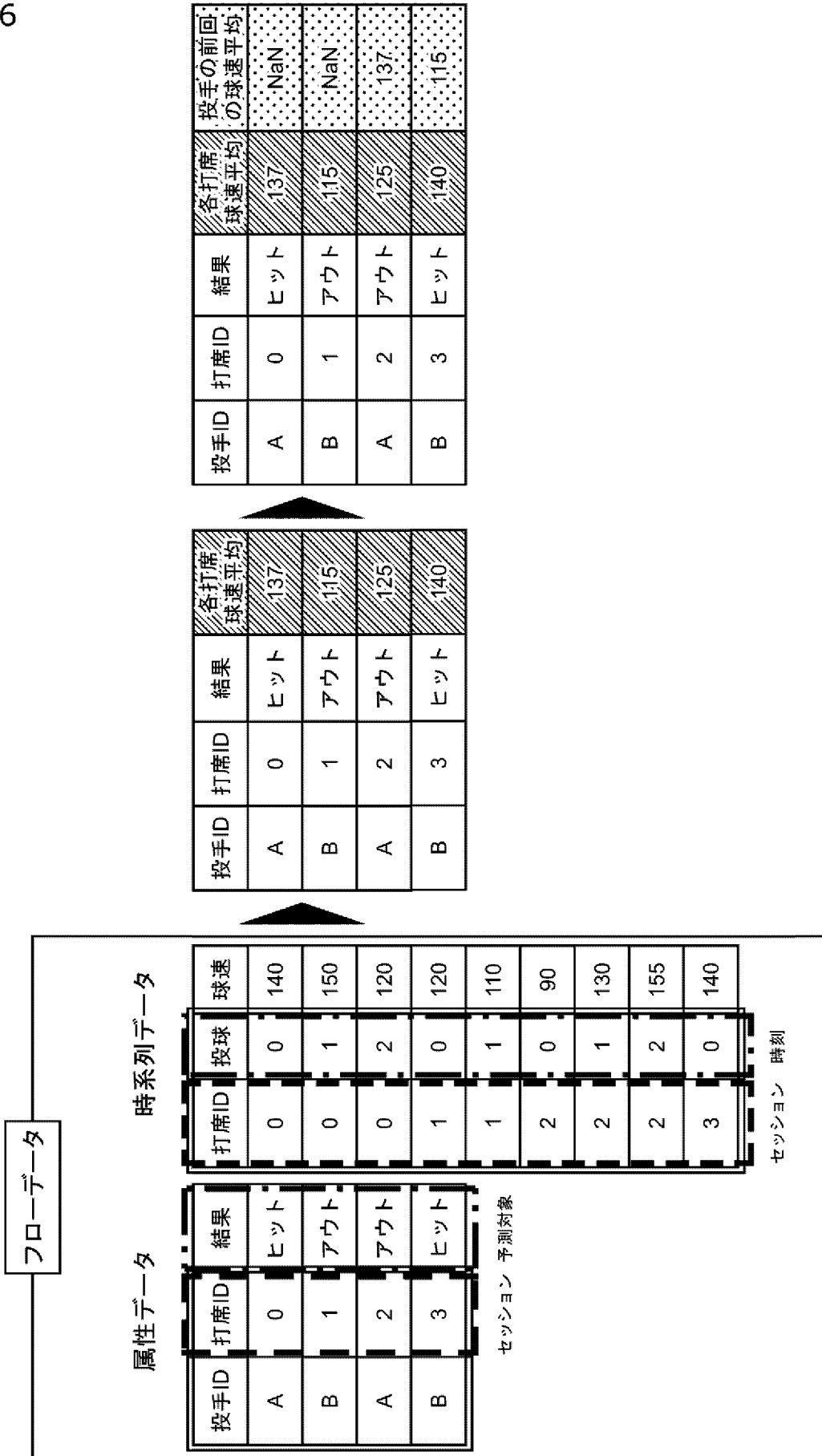


[図14]
FIG.14



[図15]
FIG.15

[図16]
FIG.16



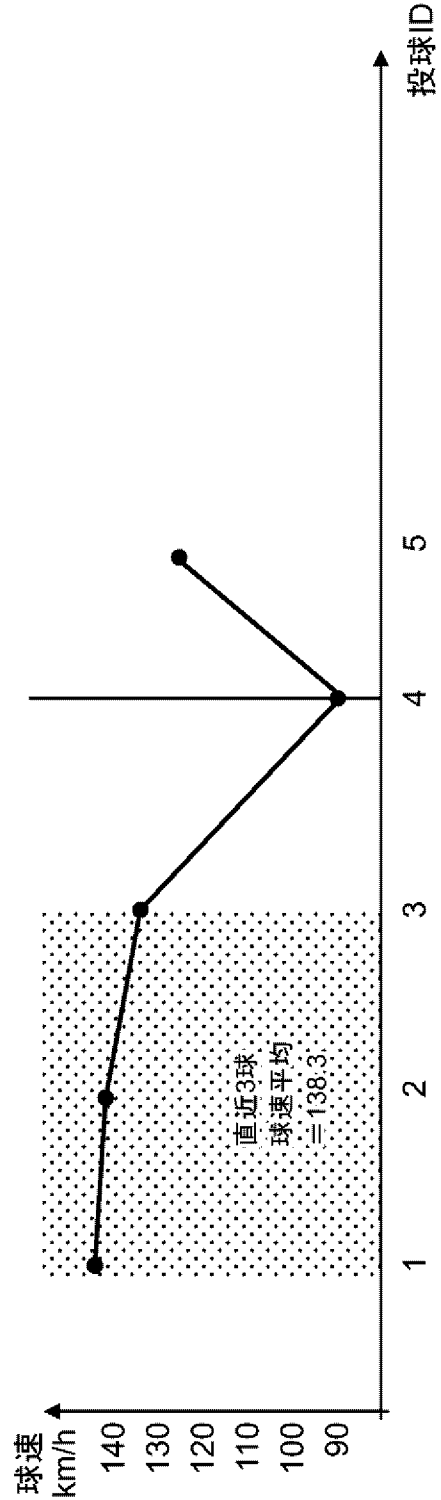
[図17]
FIG.17

フローデータ	セッションID	時刻単位	属性データ	時系列データ	セッション内特徴量	セッション間特徴量	セッション集合ID
病院バイタルログ	患者ID	日時	患者の性別	1分あたり心拍	患者の平均心拍	病院別患者の年齢	病院ID
工場ロボット稼働	稼働ID	日時	ロボットの設置場所	その日のログ	その日のロボットの平均停止回数	ロボット別通算停止回数	ロボットID
野球投球ログ	打席ID	打席内球数	左/右投げ	球速	打席内平均球速	同一投手の過去3打席の結果	投手ID

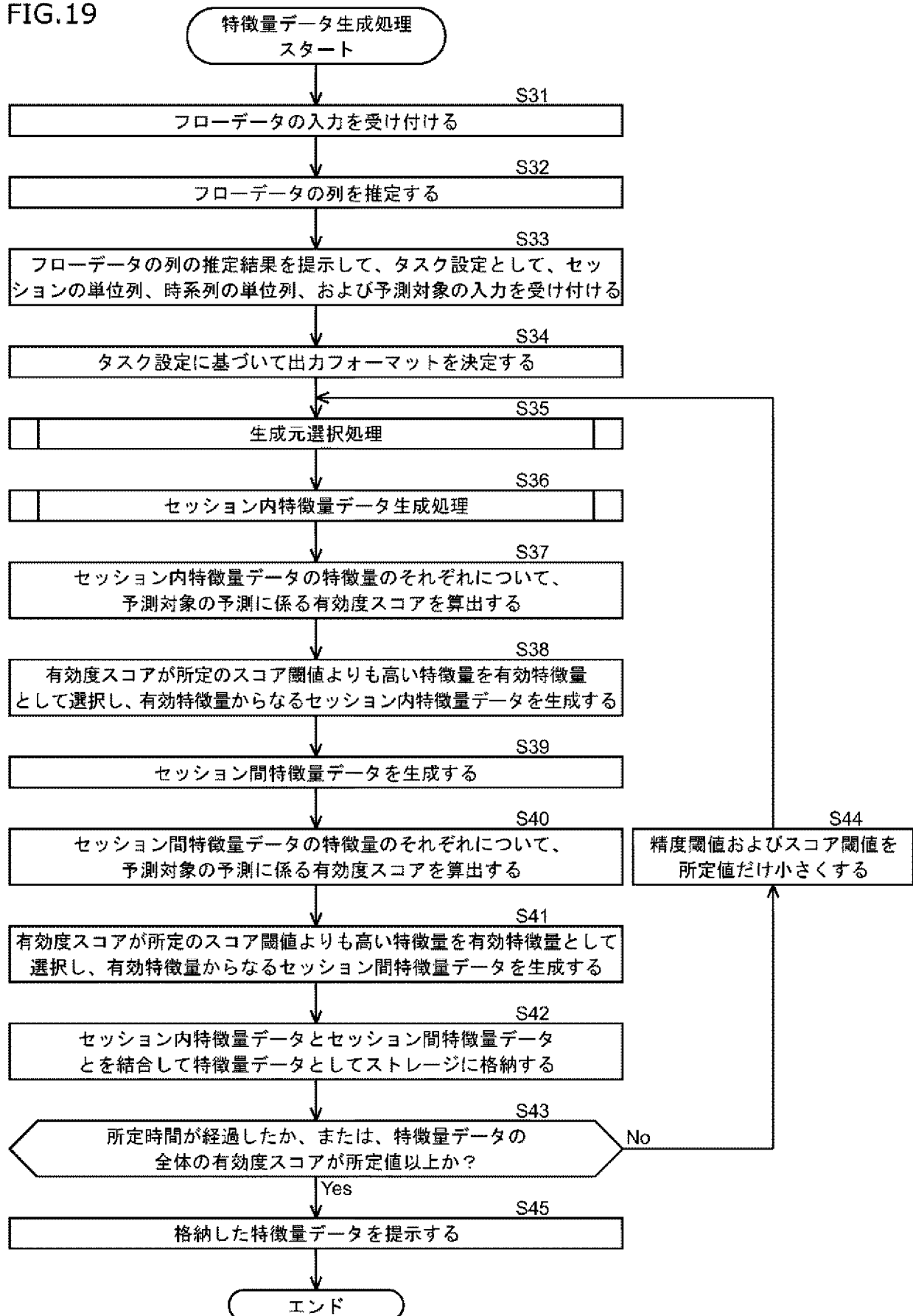
[図18]
FIG.18

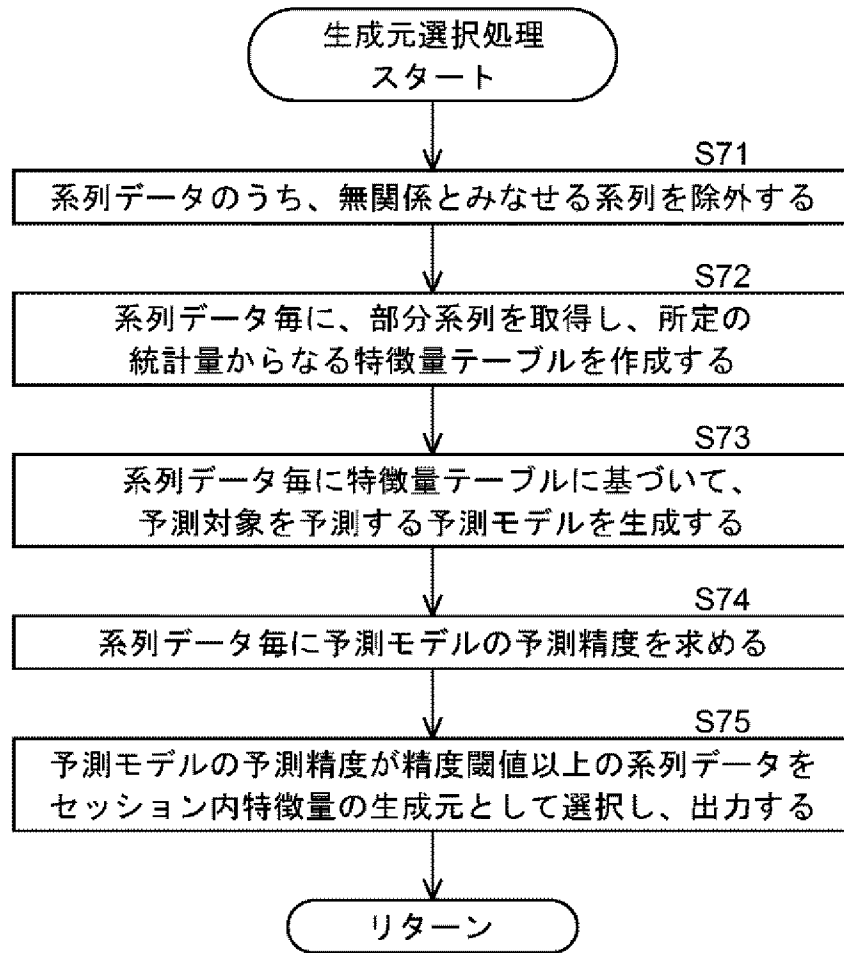
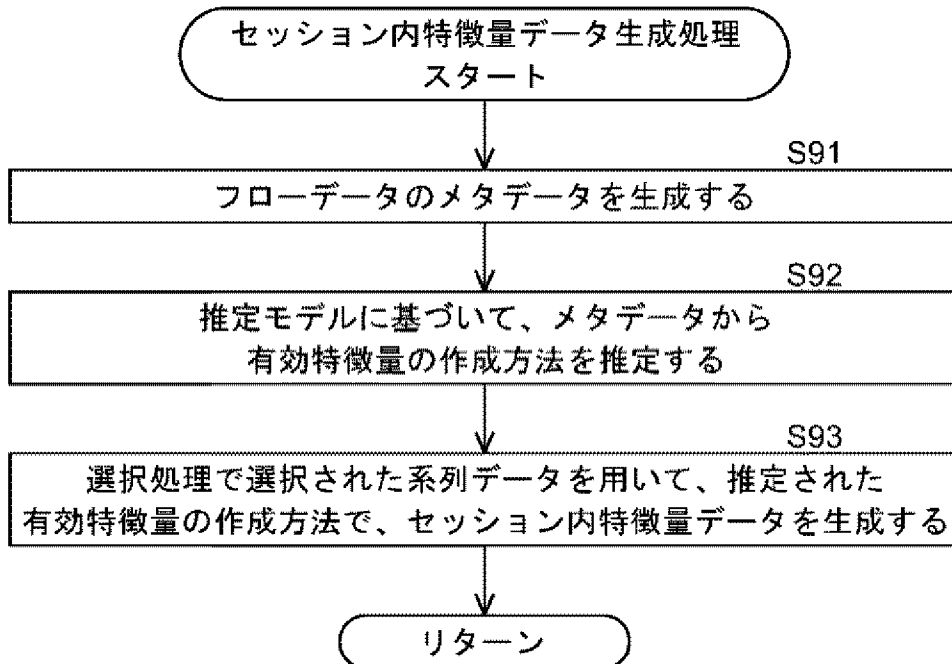
特徴量データ全体の有効度スコア : 85/100

セッションID	セッション属性データ		時刻	予測対象	セッション内特徴量データ			セッション間特徴量データ
	投手ID	結果			1球前の球速 (km/h)	直近3球球速平均(km/h)	1球前の球種	
1	A	ヒット	1	球速 (km/h)	143.9	NaN	NaN	120.4
2	A	ヒット	2	球速 (km/h)	140.2	143.9	ストレート	120.4
3	A	ヒット	3	球速 (km/h)	130.9	140.2	スライダー	120.4
4	A	ヒット	4	球速 (km/h)	90.4	130.9	チェンジアップ	120.4
5	A	ヒット	5	球速 (km/h)	124.3	90.4	スローボール	120.4
6	B	アウト	1	球速 (km/h)	150.2	NaN	NaN	144.2

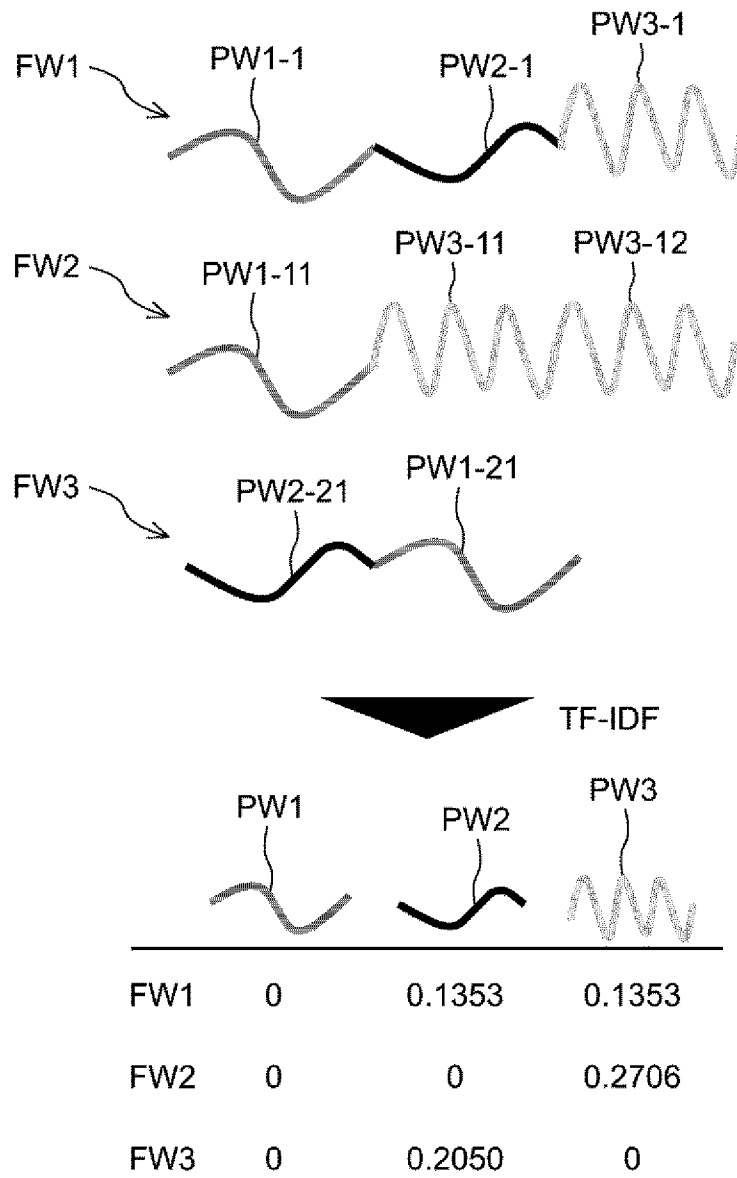


[図19]
FIG.19

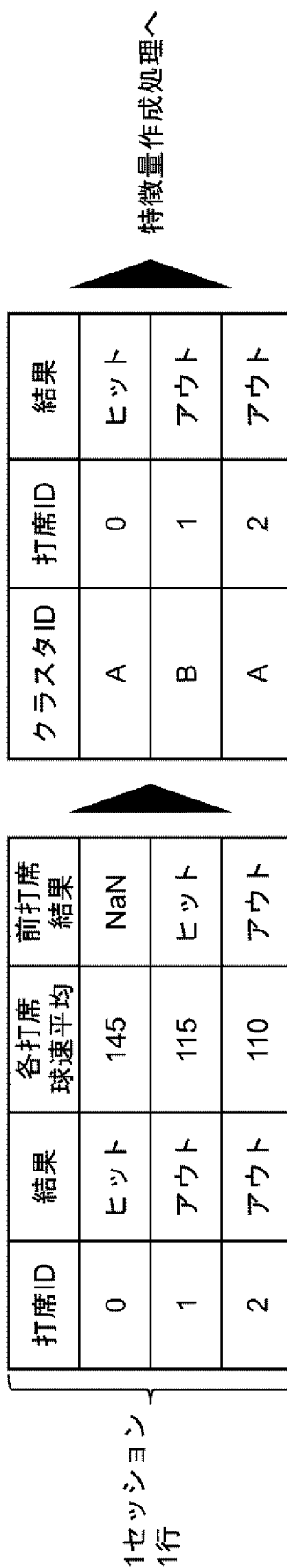


[図20]
FIG.20[図21]
FIG.21

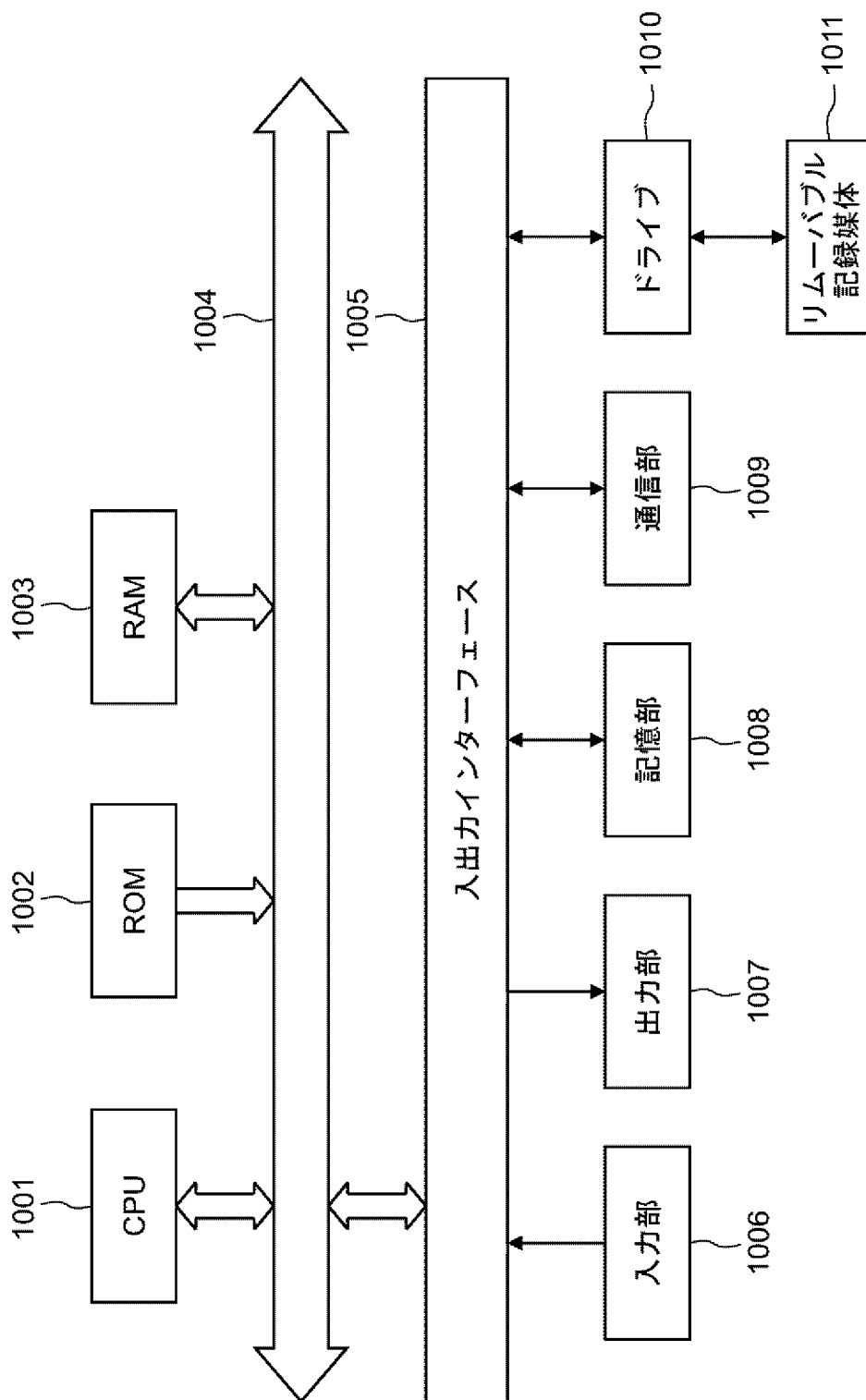
[図22]
FIG.22



[図23]
FIG.23



[図24]
FIG.24



INTERNATIONAL SEARCH REPORT

International application No.

PCT/JP2023/029935

A. CLASSIFICATION OF SUBJECT MATTER		
G06N 20/00(2019.01)i FI: G06N20/00		
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) G06N20/00		
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 1971-2023 Registered utility model specifications of Japan 1996-2023 Published registered utility model applications of Japan 1994-2023		
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.
A	JP 2019-185751 A (HITACHI LTD) 24 October 2019 (2019-10-24) entire text, all drawings	1-17
A	US 2022/0101190 A1 (ALTERYX, INC.) 31 March 2022 (2022-03-31) entire text, all drawings	1-17
A	JP 2021-60692 A (TOSHIBA CORP) 15 April 2021 (2021-04-15) entire text, all drawings	1-17
P, A	JP 2023-61486 A (MITSUBISHI HEAVY IND LTD) 02 May 2023 (2023-05-02) entire text, all drawings	1-17
<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 24 October 2023		Date of mailing of the international search report 07 November 2023
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/JP2023/029935

Patent document cited in search report			Publication date (day/month/year)	Patent family member(s)	Publication date (day/month/year)
JP	2019-185751	A	24 October 2019	US 2019/0304603 A1 entire text, all drawings	
US	2022/0101190	A1	31 March 2022	KR 10-2023-0078764 A entire text, all drawings	
				JP 2023-544011 A entire text, all drawings	
JP	2021-60692	A	15 April 2021	(Family: none)	
JP	2023-61486	A	02 May 2023	(Family: none)	

A. 発明の属する分野の分類（国際特許分類（IPC）） G06N 20/00(2019.01)i FI: G06N20/00		
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）） G06N20/00 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 - 2023年 日本国実用新案登録公報 1996 - 2023年 日本国登録実用新案公報 1994 - 2023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
A	JP 2019-185751 A (株式会社日立製作所) 24.10.2019 (2019 - 10 - 24) 全文,全図	1-17
A	US 2022/0101190 A1 (Alteryx, Inc.) 31.03.2022 (2022 - 03 - 31) 全文,全図	1-17
A	JP 2021-60692 A (株式会社東芝) 15.04.2021 (2021 - 04 - 15) 全文,全図	1-17
P, A	JP 2023-61486 A (三菱重工業株式会社) 02.05.2023 (2023 - 05 - 02) 全文,全図	1-17
<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、一般的技术水準を示すもの “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	24.10.2023	国際調査報告の発送日 07.11.2023
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） 石坂 知樹 5B 5378 電話番号 03-3581-1101 内線 3545	

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/JP2023/029935

引用文献	公表日	パテントファミリー文献	公表日
JP 2019-185751 A	24.10.2019	US 2019/0304603 A1 全文,全図	
US 2022/0101190 A1	31.03.2022	KR 10-2023-0078764 A 全文,全図	
		JP 2023-544011 A 全文,全図	
JP 2021-60692 A	15.04.2021	(ファミリーなし)	
JP 2023-61486 A	02.05.2023	(ファミリーなし)	