

(19)日本国特許庁(JP)

(12)特許公報(B2)

(11)特許番号
特許第7680980号
(P7680980)

(45)発行日 令和7年5月21日(2025.5.21)

(24)登録日 令和7年5月13日(2025.5.13)

(51)国際特許分類

F I

H 0 4 L 41/142 (2022.01) H 0 4 L 41/142

H 0 4 L 43/062 (2022.01) H 0 4 L 43/062

請求項の数 8 (全10頁)

(21)出願番号	特願2022-51846(P2022-51846)	(73)特許権者	000208891
(22)出願日	令和4年3月28日(2022.3.28)		K D D I 株式会社
(65)公開番号	特開2023-144725(P2023-144725 A)		東京都新宿区西新宿二丁目 3 番 2 号
(43)公開日	令和5年10月11日(2023.10.11)	(74)代理人	100106002
審査請求日	令和6年2月9日(2024.2.9)		弁理士 正林 真之
		(74)代理人	100120891
			弁理士 林 一好
		(72)発明者	奥井 宣広
			埼玉県ふじみ野市大原二丁目 1 番 1 5 号
			株式会社 K D D I 総合研究所内
		(72)発明者	中原 正隆
			埼玉県ふじみ野市大原二丁目 1 番 1 5 号
			株式会社 K D D I 総合研究所内
		(72)発明者	窪田 歩
			埼玉県ふじみ野市大原二丁目 1 番 1 5 号
			最終頁に続く

(54)【発明の名称】 複雑性算出装置、複雑性算出方法及び複雑性算出プログラム

(57)【特許請求の範囲】

【請求項 1】

複数のフィールドを持つ連続データに対して、特徴量として1つ以上のフィールドを選
定する特徴量選定部と、

連続する所定のデータ数の特徴量からなる特徴量セットを順に複数抽出する特徴量セッ
ト抽出部と、

複数抽出された前記特徴量セットの中から、一意な特徴量セットの集合を作成する集合
作成部と、

前記一意な特徴量セットの集合において、各特徴量セットのエントロピーを算出し、当
該エントロピーの総和を計算するエントロピー算出部と、

前記エントロピーの総和を所定の値域に正規化し、複雑性指標として出力する指標出力
部と、を備え、

前記特徴量選定部は、複数のパターンで特徴量を選定し、

前記指標出力部は、前記複数のパターンのそれぞれから算出された複数の前記複雑性指標
から統計量を求め、出力値とする複雑性算出装置。

【請求項 2】

複数のフィールドを持つ連続データに対して、特徴量として1つ以上のフィールドを選
定する特徴量選定部と、

連続する所定のデータ数の特徴量からなる特徴量セットを順に複数抽出する特徴量セッ
ト抽出部と、

複数抽出された前記特徴量セットの中から、一意な特徴量セットの集合を作成する集合作成部と、

前記一意な特徴量セットの集合において、各特徴量セットのエントロピーを算出し、当該エントロピーの総和を計算するエントロピー算出部と、

前記エントロピーの総和を所定の値域に正規化し、複雑性指標として出力する指標出力部と、を備え、

前記特徴量セット抽出部は、前記データ数を複数設定し、

前記指標出力部は、複数の前記データ数のそれぞれから算出された複数の前記複雑性指標から統計量を求め、出力値とする複雑性算出装置。

【請求項 3】

10

前記連続データは、通信情報の時系列データである請求項 1 又は請求項 2 に記載の複雑性算出装置。

【請求項 4】

前記時系列データは、フローデータである請求項 3 に記載の複雑性算出装置。

【請求項 5】

前記特徴量選定部は、複数のフィールド間で特定の処理を行い、新たに作成したフィールドを前記特徴量の一部として選定する請求項 1 から請求項 4 のいずれかに記載の複雑性算出装置。

【請求項 6】

複数のフィールドを持つ連続データに対して、特徴量として 1 つ以上のフィールドを選定する特徴量選定ステップと、

20

連続する所定のデータ数の特徴量からなる特徴量セットを順に複数抽出する特徴量セット抽出ステップと、

複数抽出された前記特徴量セットの中から、一意な特徴量セットの集合を作成する集合作成ステップと、

前記一意な特徴量セットの集合において、各特徴量セットのエントロピーを算出し、当該エントロピーの総和を計算するエントロピー算出ステップと、

前記エントロピーの総和を所定の値域に正規化し、複雑性指標として出力する指標出力ステップと、をコンピュータが実行し、

前記特徴量選定ステップにおいて、複数のパターンで特徴量を選定し、

30

前記指標出力ステップにおいて、前記複数のパターンのそれぞれから算出された複数の前記複雑性指標から統計量を求め、出力値とする複雑性算出方法。

【請求項 7】

複数のフィールドを持つ連続データに対して、特徴量として 1 つ以上のフィールドを選定する特徴量選定ステップと、

連続する所定のデータ数の特徴量からなる特徴量セットを順に複数抽出する特徴量セット抽出ステップと、

複数抽出された前記特徴量セットの中から、一意な特徴量セットの集合を作成する集合作成ステップと、

前記一意な特徴量セットの集合において、各特徴量セットのエントロピーを算出し、当該エントロピーの総和を計算するエントロピー算出ステップと、

40

前記エントロピーの総和を所定の値域に正規化し、複雑性指標として出力する指標出力ステップと、をコンピュータが実行し、

前記特徴量セット抽出ステップにおいて、前記データ数を複数設定し、

前記指標出力ステップにおいて、複数の前記データ数のそれぞれから算出された複数の前記複雑性指標から統計量を求め、出力値とする複雑性算出方法。

【請求項 8】

請求項 1 から請求項 5 のいずれかに記載の複雑性算出装置としてコンピュータを機能させるための複雑性算出プログラム。

【発明の詳細な説明】

50

【技術分野】

【0001】

本発明は、通信データの複雑性を算出するための装置、方法及びプログラムに関する。

【背景技術】

【0002】

従来、異常検知等の通信データを分析する手法を選定するため、あるいは、通信デバイスの特性又は機種等の推測又はグルーピングのために、通信データの複雑性を評価して利用することが行われている。

【0003】

例えば、特許文献1では、デバイス識別において、パケットシーケンス、パケットサイズ、パケットの間隔を含む複数のパケット間の相関、又はエントロピーに基づいて、デバイス検出プロセスを実行することが示されている。

ここで算出されるエントロピーは、個々のレコードに対して算出されるものである。

【先行技術文献】

【特許文献】

【0004】

【文献】特表2020-518208号公報

【発明の概要】

【発明が解決しようとする課題】

【0005】

時系列データに対する従来の複雑性指標は、センサデータ等の1次元を対象としたものであるため、複数の次元を持つ通信データに対して、1次元の手法をそのまま適用して複雑性を評価することは難しかった。

また、通信データの複雑性指標としてエントロピーを算出する手法はあるが、1つのレコード内の特定のフィールドに着目して計算するものであるため、個々のレコードの複雑性を評価することはできるが、通信の連続性を考慮した複雑性を評価することができなかった。

【0006】

本発明は、データの連続性を考慮した複雑性を評価できる複雑性算出装置、複雑性算出方法及び複雑性算出プログラムを提供することを目的とする。

【課題を解決するための手段】

【0007】

本発明に係る複雑性算出装置は、複数のフィールドを持つ連続データに対して、特徴量として1つ以上のフィールドを選定する特徴量選定部と、連続する所定のデータ数の特徴量からなる特徴量セットを順に複数抽出する特徴量セット抽出部と、複数抽出された前記特徴量セットの中から、ユニークな特徴量セットの集合を作成する集合作成部と、前記ユニークな特徴量セットの集合において、各特徴量セットのエントロピーを算出し、当該エントロピーの総和を計算するエントロピー算出部と、前記エントロピーの総和を所定の値域に正規化し、複雑性指標として出力する指標出力部と、を備える。

【0008】

前記連続データは、通信情報の時系列データであってもよい。

【0009】

前記時系列データは、フローデータであってもよい。

【0010】

前記特徴量選定部は、複数のフィールド間で特定の処理を行い、新たに作成したフィールドを前記特徴量の一部として選定してもよい。

【0011】

前記特徴量選定部は、複数のパターンで特徴量を選定し、前記指標出力部は、前記複数のパターンのそれぞれから算出された複数の前記複雑性指標から統計量を求め、出力値としてもよい。

10

20

30

40

50

【 0 0 1 2 】

前記特徴量セット抽出部は、前記データ数を複数設定し、前記指標出力部は、複数の前記データ数のそれぞれから算出された複数の前記複雑性指標から統計量を求め、出力値としてもよい。

【 0 0 1 3 】

本発明に係る複雑性算出方法は、複数のフィールドを持つ連続データに対して、特徴量として1つ以上のフィールドを選定する特徴量選定ステップと、連続する所定のデータ数の特徴量からなる特徴量セットを順に複数抽出する特徴量セット抽出ステップと、複数抽出された前記特徴量セットの中から、ユニークな特徴量セットの集合を作成する集合作成ステップと、前記ユニークな特徴量セットの集合において、各特徴量セットのエントロピーを算出し、当該エントロピーの総和を計算するエントロピー算出ステップと、前記エントロピーの総和を所定の値域に正規化し、複雑性指標として出力する指標出力ステップと、をコンピュータが実行する。

10

【 0 0 1 4 】

本発明に係る複雑性算出プログラムは、前記複雑性算出装置としてコンピュータを機能させるためのものである。

【 発明の効果 】

【 0 0 1 5 】

本発明によれば、データの連続性を考慮した複雑性を評価することができる。

【 図面の簡単な説明 】

20

【 0 0 1 6 】

【 図 1 】 実施形態における複雑性算出装置の機能構成を示す図である。

【 図 2 】 実施形態における複雑性算出方法の流れを示すフローチャートである。

【 図 3 】 実施形態における連続データとして取得した I P F I X レコードを例示する図である。

【 発明を実施するための形態 】

【 0 0 1 7 】

以下、本発明の実施形態の一例について説明する。

本実施形態の複雑性算出方法では、I P F I X 等の集約情報、又は通信データのヘッダ情報を使用して、デバイスが送受信する通信において、通信の連続性を考慮した複雑性を示す指標が出力される。

30

【 0 0 1 8 】

図 1 は、本実施形態における複雑性算出装置 1 の機能構成を示す図である。

複雑性算出装置 1 は、制御部 1 0 及び記憶部 2 0 の他、各種データの入出力デバイス及び通信デバイス等を備えた情報処理装置である。

【 0 0 1 9 】

制御部 1 0 は、複雑性算出装置 1 の全体を制御する部分であり、記憶部 2 0 に記憶された各種プログラムを適宜読み出して実行することにより、本実施形態における各機能を実現する。制御部 1 0 は、C P U であってよい。

具体的には、制御部 1 0 は、特徴量選定部 1 1 と、特徴量セット抽出部 1 2 と、集合作成部 1 3 と、エントロピー算出部 1 4 と、指標出力部 1 5 とを備える。

40

【 0 0 2 0 】

記憶部 2 0 は、ハードウェア群を複雑性算出装置 1 として機能させるための各種プログラム、及び各種データ等の記憶領域であり、R O M、R A M、フラッシュメモリ又はハードディスクドライブ (H D D) 等であってよい。具体的には、記憶部 2 0 は、本実施形態の各機能を制御部 1 0 に実行させるためのプログラム (複雑性算出プログラム) を記憶し、さらに、処理対象として受信した連続データ、及び演算結果等を記憶する。

【 0 0 2 1 】

特徴量選定部 1 1 は、ユーザからの選択入力等に従って、複数のフィールドを持つ連続データに対して、特徴量として1つ以上のフィールドを選定する。

50

さらに、特徴量選定部 11 は、複数のフィールド間で四則演算等、特定の処理を行い、新たに作成したフィールドを特徴量の一部として選定してもよい。

また、特徴量選定部 11 は、複数のパターンで特徴量を選定してもよい。

【0022】

ここで、連続データは、例えば、IoT デバイス等からの出力データの他、双方向の通信情報の時系列データであってもよい。通信情報は、例えばパケットのヘッダ情報であってもよいし、統計情報が格納されたフローデータの列であってもよい。

また、これらの連続データは、例えば、固定の周期でサンプリングされたものであってもよい。

【0023】

特徴量セット抽出部 12 は、連続する所定のデータ数の特徴量からなる特徴量セットを順に複数抽出する。

このとき、特徴量セット抽出部 12 は、データ数を複数設定してもよい。この場合、特徴量セット抽出部 12 は、データ数毎に特徴量セットを抽出する。

【0024】

集合作成部 13 は、複数抽出された特徴量セットの中から、ユニークな特徴量セットの集合を作成する。

【0025】

エントロピー算出部 14 は、ユニークな特徴量セットの集合において、各特徴量セットのエントロピーを算出し、これらのエントロピーの総和を計算する。

【0026】

指標出力部 15 は、計算されたエントロピーの総和を所定の値域に正規化し、複雑性指標として出力する。

このとき、指標出力部 15 は、複数のパターンで選定された特徴量、及び複数設定されたデータ数のそれぞれから算出された複数の複雑性指標から統計量を求め、最終的な複雑性指標として出力してもよい。

統計量は、例えば、平均値、中央値、最大値、最小値等であってよい。また、指標出力部 15 は、複数の複雑性指標から標準偏差を計算し、複雑性指標に幅を持たせて出力してもよい。

【0027】

図 2 は、本実施形態における複雑性算出方法の流れを示すフローチャートである。

ここでは、通信データとして IPFIX を用いた場合の具体的な複雑性指標の算出手順を例示する。

【0028】

IPFIX には、次のようなフィールド (IPFIX エlement) がある。

flowEndMilliseconds : フローが終了した時刻。

flowStartMilliseconds : フローが開始した時刻。

flowDurationMilliseconds : フローの継続時間。

reverseFlowDeltaMilliseconds : フローの最初の行きパケットと帰りパケットとの時間の差。

protocolIdentifier : フローのプロトコル (ICMP、TCP、UDP 等)。

sourceIPv4Address : 送信元の IPv4 アドレス。

sourceTransportPort : 送信元のポート番号。

packetTotalCount : フローに含まれる行き方向のパケット数。

octetTotalCount : フローに含まれる行き方向のオクテット数。

sourceMacAddress : 送信元の MAC アドレス。

destinationIPv4Address : 宛先の IPv4 アドレス。

destinationTransportPort : 宛先のポート番号。

reversePacketTotalCount : フローに含まれる帰り方向のパケット数。

reverseOctetTotalCount : フローに含まれる帰り方向のオクテット数。

10

20

30

40

50

【 0 0 2 9 】

ステップ S 1 において、特徴量選定部 1 1 は、これらのフィールドから c 個のフィールドを選定し、c 個のフィールドの組み合わせを特徴量 F とする。例えば、3 つフィールドを選定した場合、i 番目のレコードの特徴量は、

$F_i = (\text{packetTotalCount}, \text{octetTotalCount}, \text{reversePacketTotalCount})$ のようになる。

また、前述のように、特徴量選定部 1 1 は、これらのフィールドのいずれかを用いて、例えば、送受信比率（送信パケット数 / 受信パケット数）等の演算結果を特徴量に含めてもよい。

【 0 0 3 0 】

10

図 3 は、本実施形態における連続データとして取得した I P F I X レコードを例示する図である。

ここでは、選定された 3 つのフィールドのみを示しており、複雑性を計測する期間として取得したレコード総数は 8（レコード番号 1 ~ 8）である。

【 0 0 3 1 】

ステップ S 2 において、特徴量セット抽出部 1 2 は、d 個の連続データを 1 つのセットとして扱い、各セットから特徴量セットを抽出する。

例えば、図 3 のような I P F I X レコードにおいて、d = 3 とした場合、特徴量セットとなりうるレコード番号の組み合わせは次のようになる。

$\{(1,2,3), (2,3,4), (3,4,5), (4,5,6), (5,6,7), (6,7,8)\}$

20

なお、複雑性指標の算出に使用するレコード数の総数を z（図 3 の例では 8）とすると、特徴量セットの数は、 $z - d + 1$ （ $= 8 - 3 + 1 = 6$ ）となる。

【 0 0 3 2 】

ここで、抽出される特徴量セットは、例えば、レコード番号が（1, 2, 3）の組み合わせの場合、

$F_{(1,2,3)} = \{F_1, F_2, F_3\} = \{(8, 400, 300), (8, 400, 300), (4, 160, 80)\}$ である。

【 0 0 3 3 】

ステップ S 3 において、集合作成部 1 3 は、生成された特徴量セットの中から、ユニークな特徴量セットの集合を作成する。

30

図 3 の例の場合、 $F_{(1,2,3)} = F_{(5,6,7)}$ となるため、ユニークな特徴量セットの集合は、 $\{(1,2,3), (2,3,4), (3,4,5), (4,5,6), (6,7,8)\}$ となる。

【 0 0 3 4 】

ステップ S 4 において、エントロピー算出部 1 4 は、複雑性を計測する期間に含まれる通信データにおいて、各特徴量セットの出現頻度（出現数）を計算する。

例えば、図 3 の例の場合、特徴量セット毎の出現頻度は、次のように計算される。

$(1,2,3)=2, (2,3,4)=1, (3,4,5)=1, (4,5,6)=1, (6,7,8)=1$

【 0 0 3 5 】

ステップ S 5 において、エントロピー算出部 1 4 は、ユニークな特徴量セットの集合における要素毎のエントロピー（平均情報量）を計算し、さらに、各特徴量セットのエントロピーを総和した合計値を計算する。

40

【 0 0 3 6 】

ここで、ユニークな特徴量セットの集合における要素数を N とし、i 番目の要素の出現頻度を V_i とすると、各特徴量セットのエントロピーの総和は、

【数 1】

$$E = - \sum_{i=1}^N \frac{V_i}{z - d + 1} \ln \frac{V_i}{z - d + 1}$$

50

であり、図 3 の例では、

【数 2】

$$E = -\left(\frac{2}{6} \ln \frac{2}{6}\right) - \left(\frac{1}{6} \ln \frac{1}{6}\right) \times 4 \cong 0.3662 + 1.1945 = 1.5607$$

となる。

【0037】

ステップ S 6 において、指標出力部 15 は、エントロピーの合計値 E を、要素数 N に応じた最大エントロピーで割り、値の範囲が 0 以上 1 以下となるように正規化して複雑性指標 E_p とする。

10

【数 3】

$$E_p = \frac{E}{\ln N}$$

図 3 の例では、複雑性指標 E_p は、

【数 4】

$$E_p = \frac{1.5607}{1.6094} \cong 0.9697$$

20

となる。

【0038】

なお、前述のように、複数のパターンでフィールド選定された場合、又は特徴量セットに含める連続データ数 d が複数設定された場合、パターン及びデータ数毎に、ステップ S 1 ~ S 6 が実行され、複数の複雑性指標が算出される。

【0039】

ステップ S 7 において、指標出力部 15 は、算出した複雑性指標 E_p を出力する。複数の複雑性指標 E_p が算出された場合、指標出力部 15 は、これらの統計量を求め、最終的な複雑性指標として出力する。

30

【0040】

本実施形態によれば、複雑性算出装置 1 は、複数のフィールドを持つ連続データに対して、連続する所定のデータ数の特徴量からなる特徴量セットを順に複数抽出し、ユニークな特徴量セットのエントロピーを総和することにより複雑性指標を算出する。

したがって、複雑性算出装置 1 は、センサの出力や通信の時系列データ等から取得した連続データに対して、データの連続性を考慮した複雑性を、処理負荷の小さい簡易な計算により評価することができる。

【0041】

この結果、例えば、通信データに対する異常検知を実施する場合に、通信データの複雑性を事前に把握することで、複雑性が低い場合にはホワイトリストを、複雑性が高い場合には決定木、あるいはディープラーニングを用いる等、分析手法の選定範囲を絞ることができる。

40

また、通信データの複雑性によって、通信元のデバイスの特性又は機種等を推測し、分類することができる。

【0042】

複雑性算出装置 1 は、通信データに関して、フローデータを対象とすることにより、複数パケットが集約されたレコードから特徴量を取得できるため、連続データ数 d を小さくし、計算効率を向上できる。

【0043】

50

複雑性算出装置 1 は、複数のフィールド間で特定の処理を行い、新たに作成したフィールドを特徴量の一部として選定してもよく、これにより、多彩な特徴量を用いて、データの特性に応じた複雑性の評価を適切に実施できる。

【 0 0 4 4 】

複雑性算出装置 1 は、複数のパターンで特徴量を選定、あるいは、連続データ数 d を複数設定することで、複数の複雑性指標を算出し、統計量を出力してもよい。

これにより、複雑性算出装置 1 は、最適な特徴量及びデータ数が不明な場合であっても、評価結果の信頼性を向上できる他、最大値又は最小値等の要求された指標を適切に出力できる。

なお、連続データ数 d は、2 以上 7 以下程度とするのが好ましく、これにより、計算量を抑えつつ、連続データの特徴を適切に抽出して複雑性を評価できる。

10

【 0 0 4 5 】

前述の実施形態により、例えば、通信データの分析を効率化できることから、国連が主導する持続可能な開発目標 (SDGs) の目標 9「レジリエントなインフラを整備し、持続可能な産業化を推進するとともに、イノベーションの拡大を図る」に貢献することが可能となる。

【 0 0 4 6 】

以上、本発明の実施形態について説明したが、本発明は前述した実施形態に限るものではない。また、前述した実施形態に記載された効果は、本発明から生じる最も好適な効果を列挙したに過ぎず、本発明による効果は、実施形態に記載されたものに限定されるものではない。

20

【 0 0 4 7 】

複雑性算出装置 1 による複雑性算出方法は、ソフトウェアにより実現される。ソフトウェアによって実現される場合には、このソフトウェアを構成するプログラムが、情報処理装置 (コンピュータ) にインストールされる。また、これらのプログラムは、CD-ROM のようなリムーバブルメディアに記録されてユーザに配布されてもよいし、ネットワークを介してユーザのコンピュータにダウンロードされることにより配布されてもよい。さらに、これらのプログラムは、ダウンロードされることなくネットワークを介した Web サービスとしてユーザのコンピュータに提供されてもよい。

【 符号の説明 】

30

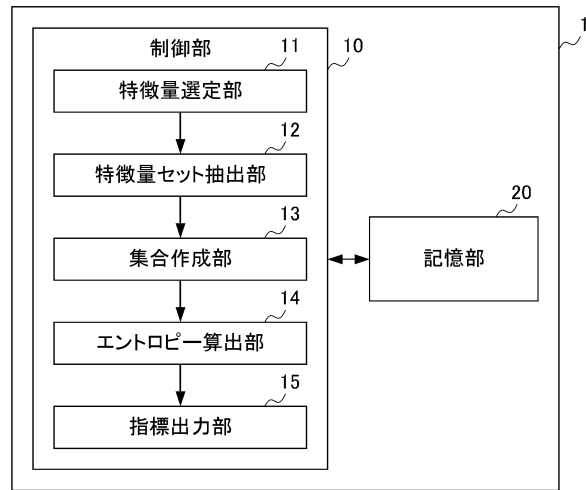
【 0 0 4 8 】

- 1 複雑性算出装置
- 10 制御部
- 11 特徴量選定部
- 12 特徴量セット抽出部
- 13 集合作成部
- 14 エントロピー算出部
- 15 指標出力部
- 20 記憶部

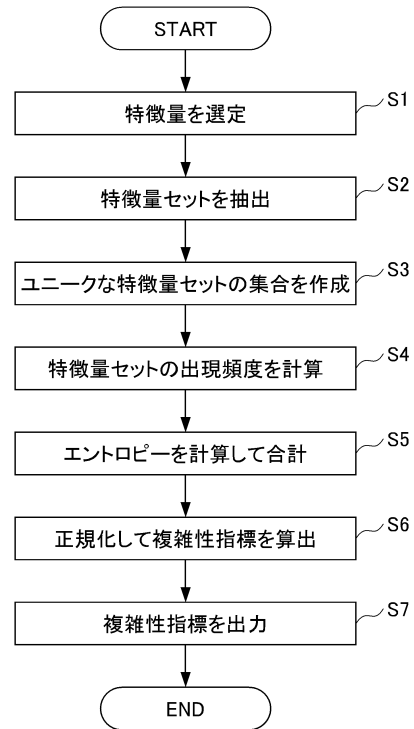
40

【図面】

【図 1】



【図 2】



【図 3】

No.	packetTotalCount	octetTotalCount	reversePacketTotalCount
1	8	400	300
2	8	400	300
3	4	160	80
4	4	160	80
5	8	400	300
6	8	400	300
7	4	160	80
8	4	160	120

10

20

30

40

50

フロントページの続き

株式会社 K D D I 総合研究所内

審査官 長濱 美紗

- (56)参考文献 国際公開第 2 0 1 9 / 1 7 6 9 9 7 (W O , A 1)
特開 2 0 1 9 - 1 4 5 0 8 1 (J P , A)
米国特許出願公開第 2 0 1 6 / 0 2 9 2 4 1 8 (U S , A 1)
特開 2 0 1 1 - 2 4 4 0 9 8 (J P , A)
- (58)調査した分野 (Int.Cl. , D B 名)
H 0 4 L 1 2 / 0 0 - 1 2 / 6 6
4 1 / 0 0 - 1 0 1 / 6 9 5