

(19)日本国特許庁(JP)

**(12)特許公報(B2)**

(11)特許番号  
**特許第6992733号**  
**(P6992733)**

(45)発行日 令和4年1月13日(2022.1.13)

(24)登録日 令和3年12月13日(2021.12.13)

(51)国際特許分類

G 0 6 F	3/12 (2006.01)	F I	G 0 6 F	3/12	3 4 4
B 4 1 J	29/38 (2006.01)		G 0 6 F	3/12	3 0 3
G 0 6 N	20/00 (2019.01)		G 0 6 F	3/12	3 7 8
			G 0 6 F	3/12	3 5 1
			G 0 6 F	3/12	3 4 3

請求項の数 8 (全18頁) 最終頁に続く

(21)出願番号 特願2018-218996(P2018-218996)  
 (22)出願日 平成30年11月22日(2018.11.22)  
 (65)公開番号 特開2020-86814(P2020-86814A)  
 (43)公開日 令和2年6月4日(2020.6.4)  
 審査請求日 令和2年6月8日(2020.6.8)

(73)特許権者 000002369  
 セイコーエプソン株式会社  
 東京都新宿区新宿四丁目1番6号  
 (74)代理人 100179475  
 弁理士 仲井 智至  
 100216253  
 弁理士 松岡 宏紀  
 100225901  
 弁理士 今村 真之  
 (72)発明者 大宮 英久  
 長野県諏訪市大和3丁目3番5号 セイ  
 コーエプソン株式会社内  
 審査官 白石 圭吾

最終頁に続く

(54)【発明の名称】 情報処理装置、および情報処理方法

**(57)【特許請求の範囲】****【請求項1】**

印刷データに基づいて印刷物を印刷するレシートプリンターと通信する情報処理装置であつて、

前記レシートプリンターから送信された前記印刷データを取得する取得部と、

前記印刷データに含まれるテキストデータに基づいて、前記印刷物の種類を判定する判定部と、

印刷物の種類に対応した複数のスクリプトを記憶する記憶部と、

複数の前記スクリプトから、前記判定部で判定した前記印刷物の種類に対応した1のスクリプトを選択し、選択した前記1のスクリプトを前記テキストデータに対して適用する適用部と、

を備える情報処理装置。

**【請求項2】**

前記判定部は、

前記テキストデータと、前記印刷物の前記種類とを対応づけた教師データに基づいて機械学習された学習モデルにより、前記種類の判定を行う、

請求項1に記載の情報処理装置。

**【請求項3】**

前記判定部は、

前記印刷物が複数の前記種類のそれぞれである可能性を取得し、

前記適用部は、

前記可能性が高い前記種類であるほど順位が高くなるように、複数の前記スクリプトから前記 1 のスクリプトを選択する優先順位を設定する、  
請求項 1 または請求項 2 に記載の情報処理装置。

【請求項 4】

前記適用部は、

選択した前記 1 のスクリプトを前記テキストデータに対して適用し、前記 1 のスクリプトを適用した後に、設定した前記優先順位に従って、前記 1 のスクリプトと異なる他のスクリプトを選択して適用する、

請求項 3 に記載の情報処理装置。

10

【請求項 5】

前記印刷物の前記種類には、店舗での購入記録を示す前記印刷物と、店舗の売上集計を示す前記印刷物と、が含まれる、

請求項 1 ~ 請求項 4 のいずれか 1 項に記載の情報処理装置。

【請求項 6】

前記スクリプトは、

前記テキストデータに含まれる項目名と、前記項目名に対応したパラメーターとを対応づける処理を行うプログラムである、

請求項 1 ~ 請求項 5 のいずれか 1 項に記載の情報処理装置。

【請求項 7】

制御部、及び記憶部を備え、印刷データに基づいて印刷物を印刷するレシートプリンターと通信する情報処理装置の情報処理方法であって、

前記記憶部が、前記印刷物の種類に対応した複数のスクリプトを前記記憶部に記憶する記憶工程と、

前記制御部が、前記レシートプリンターから送信された前記印刷データを前記制御部で取得する取得工程と、

前記制御部が、前記印刷データに含まれるテキストデータに基づいて、前記印刷物の種類を前記制御部で判定する判定工程と、

前記制御部が、複数の前記スクリプトから、前記印刷物の種類に対応した 1 のスクリプトを選択し、選択した前記 1 のスクリプトを前記テキストデータに対して適用する適用工程と、

を含む情報処理方法。

30

【請求項 8】

前記判定工程は、

前記制御部が、前記テキストデータと、前記印刷物の前記種類とを対応づけた教師データに基づいて機械学習された学習モデルにより、前記種類の判定を行う工程である、

請求項 7 に記載の情報処理方法。

【発明の詳細な説明】

【技術分野】

【0001】

本発明は、情報処理装置、レシートプリンターおよび情報処理方法に関する。

40

【背景技術】

【0002】

従来、POS 端末等で利用されるレシートプリンターで印刷を行わせるための印刷データを解析する技術が知られている。例えば、特許文献 1 においては、印刷データを制御サーバーに送信すると、制御サーバーがレイアウトに対応する方法で印刷データを解析する技術が開示されている。

【先行技術文献】

【特許文献】

【0003】

50

【文献】特開 2015 - 158775 号公報

【発明の概要】

【発明が解決しようとする課題】

【0004】

印刷データによって印刷される印刷物のレイアウトが不明である状態では、レイアウトに対応した解析の方法が不明であるため、想定される全ての方法で解析を行い、最も適切な結果を解析結果と見なす手法が採用され得る。このように全ての方法で解析を行う構成では、解析に時間がかかるてしまう。

【課題を解決するための手段】

【0005】

上記目的を達成するための情報処理装置は、印刷データに基づいて印刷物を印刷するレシートプリンターから、印刷データを取得する取得部と、印刷データに含まれるテキストデータに基づいて、印刷物の種類を判定する判定部と、複数のスクリプトから、種類に対応したスクリプトを選択し、選択したスクリプトをテキストデータに対して適用する適用部と、を備える。すなわち、情報処理装置においては、印刷データに含まれるテキストデータに基づいて印刷物の種類に対応したスクリプトを選択する。この結果、全ての種類のスクリプトを適用しなくても、適切なスクリプトをテキストデータに対して適用できる可能性を高めることができる。

10

【0006】

また、判定部は、テキストデータと、印刷物の種類とを対応づけた教師データに基づいて機械学習された学習モデルにより、種類の判定を行う構成であってもよい。この構成によれば、人間による判断に頼らずに規定されたテキストデータと印刷物の種類との対応関係に基づいて印刷物の種類を判断することができる。

20

【0007】

さらに、判定部は、印刷物が複数の種類のそれぞれである可能性を取得し、適用部は、可能性が高い種類であるほど順位が高くなるように、印刷物の種類に対応したスクリプトを選択する優先順位を設定する構成であっても良い。この構成によれば、適切なスクリプトをテキストデータに対して適用することができる可能性を高めることができる。

【0008】

さらに、適用部は、選択したスクリプトをテキストデータに対して適用して結果が得られるまで、設定した優先順位に従って、スクリプトを選択して適用する構成であっても良い。この構成によれば、早期に適切なスクリプトをテキストデータに対して適用することができる可能性を高めることができる。

30

【0009】

さらに、印刷物の種類には、店舗での購入記録を示す印刷物と、店舗の売上集計を示す印刷物と、が含まれる構成であっても良い。この構成によれば、店舗での購入記録を示す印刷物と、店舗の売上集計を示す印刷物とが印刷され得るレシートプリンターで利用される印刷データに対して適切なスクリプトを適用することができる可能性を高めることができる。

【0010】

40

さらに、前記スクリプトは、テキストデータに含まれる項目名と、項目名に対応したパラメーターとを対応づける処理を行うためのプログラムであっても良い。この構成によれば、適切なスクリプトをテキストデータに対して適用することで、適切な解析結果を得ることができる。また、テキストデータが示す内容を容易に利用可能な状態にすることができる。

【図面の簡単な説明】

【0011】

【図1】情報処理装置のブロック図である。

【図2】ニューラルネットワークの構造を示す図である。

【図3】機械学習処理を示すフローチャートである。

50

【図4】情報処理を示すフローチャートである。

【図5】購入記録を示す印刷物と構造化データの例を示す図である。

【図6】RNNを含むモデルの構造を示す図である。

【発明を実施するための形態】

【0012】

ここでは、下記の順序に従って本発明の実施の形態について説明する。

(1) 情報処理装置の構成：

(2) 機械学習処理：

(3) 情報処理：

(4) 他の実施形態：

10

【0013】

(1) 情報処理装置の構成：

図1は、本発明の実施形態にかかる情報処理装置を示すブロック図である。本実施形態にかかる情報処理装置1は、レシートプリンター2と通信可能なサーバーであり、インターネット等のネットワークを介してレシートプリンター2と通信を行うことができる。

【0014】

レシートプリンター2は店舗のレジスターに備えられるプリンターである。すなわち、レシートプリンター2は、店舗において商品を購入した各顧客に対して購入記録を示す印刷物(レシート)を渡すために使用され、会計後にオペレーターの操作によってレシートが印刷される。もちろん、レシートプリンター2は他の機器と協働する構成であってもよく、例えば、POSシステム(point of sales system)の一部を構成していくても良い。

20

【0015】

レシートプリンター2は、図示しないレジスター等の制御装置からの指令によって各種の印刷物を印刷することができる。本実施形態において図示しない制御装置は、店舗での購入記録を示す印刷物と、店舗の売上集計を示す印刷物と、店舗で働く者の勤怠情報を示す印刷物とのいずれかを指定して印刷することができる。

【0016】

すなわち、制御装置は、図示しないプロセッサーの機能により、顧客が購入した商品の価格等を示す購入記録を示す印刷データを生成することができる。制御装置が、レシートプリンター2に対して制御信号を出力して当該印刷データに基づく印刷を指示すると、レシートプリンター2は購入記録を示す印刷物を印刷することができる。

30

【0017】

また、制御装置は、任意の期間毎の購入記録を集計して売上集計を取得することができ、当該売上集計を示す印刷データを生成することができる。制御装置が、レシートプリンター2に対して制御信号を出力して当該印刷データに基づく印刷を指示すると、レシートプリンター2は店舗の売上集計を示す印刷物を印刷することができる。さらに、制御装置は、店舗で働く者の勤務開始および終了に関する情報を記録、管理しており、任意の期間毎の情報を集計して勤務情報を取得することができる。制御装置は、当該勤務情報を示す印刷データを生成することができ、レシートプリンター2に対して制御信号を出力して当該印刷データに基づく印刷を指示すると、レシートプリンター2は店舗で働く者の勤怠情報を示す印刷物を印刷することができる。

40

【0018】

このように、レシートプリンター2は、購入記録を示す印刷物を印刷することができるが、他の印刷物である売上集計を示す印刷物、勤怠情報を示す印刷物を印刷することもできる。なお、本実施形態において、レシートプリンター2に受け渡される印刷データには、印刷物に印刷されるべきテキストを示すテキストデータが含まれている。

【0019】

本実施形態においてレシートプリンター2は、情報処理装置1に印刷データを送信可能な送信部を備える。レシートプリンター2は、印刷物の印刷指示が行われると、印刷データに基づいてレシートプリンター2の印刷ヘッド(印刷部)による印刷を行うとともに、送

50

信部の機能により、情報処理装置 1 に対して印刷データを送信する。なお、本実施形態においては、レシートプリンター 2 と情報処理装置 1 とが接続され、レシートプリンター 2 から情報処理装置 1 に印刷データが送信されるが、レシートプリンター 2 を制御する制御装置から印刷データが情報処理装置 1 に送信されてもよい。この場合、レシートプリンター 2 を制御する制御装置もレシートプリンター 2 を構成すると見なすことができる。なお、送信部は、通信インターフェース、通信回路、通信ポート等で構成される。また、送信部による印刷データの送信は、USB やイーサネット等の有線通信による送信であってもよく、Wi-Fi や Bluetooth 等の無線通信による送信であってもよい（イーサネット、Wi-Fi、Bluetooth は登録商標）。

#### 【 0 0 2 0 】

10

情報処理装置 1 は、プロセッサー 10（制御部）と、HDD（ハードディスクドライブ、記憶部）20 と、通信部 30 とを備えている。通信部 30 は、インターネット等を介して外部の機器と通信を行うための装置であり、レシートプリンター 2 と通信を行って情報を授受することができる。なお、通信部 30 は、レシートプリンター 2 の送信部と同様のハードウェア構成を有してよく、送信部と同様に、有線通信であっても、無線通信であってもよい。

#### 【 0 0 2 1 】

プロセッサー 10 は、CPU、RAM 等を備え、HDD 20 に記録された種々のプログラムを実行することができる。本実施形態においては、これらのプログラムに情報処理プログラム 11 が含まれている。情報処理プログラム 11 は、印刷データに含まれるテキストデータに対してスクリプトを適用する処理をプロセッサー 10 に実行させるプログラムである。

20

#### 【 0 0 2 2 】

本実施形態において、スクリプトは、テキストデータに含まれる特定の項目と、項目の内容を示すパラメーターとを対応づけるための処理を記述したプログラムである。例えば、図 5 の左側には購入記録を示す印刷物 Pr が示されている。当該印刷物には、印刷物が発行された日時（02 / 17 / 2017）やレジスターのオペレーター名（Name A）、商品名毎の価格（\$ 4 . 99）等が記載される。これらの日時やオペレーター名、価格は可変の情報であり本実施形態においてはこれらをパラメーターと呼ぶ。

#### 【 0 0 2 3 】

30

これらのパラメーターは、その情報自体（価格の値等）のみでは意味をなさないため、これらのパラメーターが何に対するパラメーターであるかを示す情報が項目である。例えば、4 . 99 ドルなどという値自体では何の価格であるのか不明であるため、商品名が対応づけられることで特定の商品の価格が 4 . 99 ドルであるという意味を持つ。また、例えば、2017 年 2 月 17 日などという値自体では何の日時であるのか不明であるため、印刷物が発行された日時であることが対応づけられることで、発行日時が 2017 年 2 月 17 日であるという意味を持つ。

#### 【 0 0 2 4 】

このように、本実施形態において印刷物に印刷するテキストは、パラメーター（数値または人名等の文字列）と、パラメーターが何のパラメーターであるのかを示す項目とによって構成されている。そこで、スクリプトは、テキストデータを解析し、項目とパラメーターとを抽出して両者を対応づける処理を行うように定義されている。本実施形態においては、当該処理が行われると項目名とパラメーターとを対応づけたテキストデータが生成される。ここでは、生成された当該テキストデータを構造化データと呼ぶ。なお、構造化データにおいては、項目が階層構造を有している場合がある。すなわち、ある項目がさらに下位のサブ項目に細分化されている場合がある。もちろん、階層の数は 3 以上であってもよい。図 5 に示す構造化データにおいては、括弧の種類を区別することによってこのような階層構造を示している。

40

#### 【 0 0 2 5 】

図 5 の右側には印刷物 Pr から生成された構造化データ St の例を示している。図 5 の右

50

側の構造化データ S tにおいては、項目名とパラメーターとが": "によって対応づけられている。例えば、印刷物が発行された日時を示すパラメーターである"0 2 / 1 7 / 2 0 1 7 "が" p r i n t e d \_ a t "という項目名に対応づけられている。また、レジスターを操作したオペレーター名を示すパラメーターである"Name A "が" s t a f f "という項目のサブ項目の項目名である"i d "に対応づけられている。なお、印刷物において、項目名は表記されていなくてもよい。例えば、図 5 の左側に示す印刷物 P rにおいて、印刷物が発行された日時は 0 2 / 1 7 / 2 0 1 7 という数値で表記されているが、発行日時であることを示す文言は表記されていない。

#### 【 0 0 2 6 】

本実施形態においては、上述のように 3 種類の印刷物（購入記録を示す印刷物、売上集計を示す印刷物、勤怠情報を示す印刷物）を印刷することができ、それぞれにおける項目やパラメーターは異なり得る。そこで、本実施形態においては、印刷物の種類に対応したスクリプトが予め用意され、スクリプトデータ 2 0 d として H D D 2 0 に記録される。

10

#### 【 0 0 2 7 】

情報処理装置 1 は、レシートプリンター 2 から印刷データが送信されると、当該印刷データに含まれるテキストデータに対してスクリプトを適用するが、印刷データによって印刷される印刷物の種類が不明であると、適用するスクリプトを決定することができない。印刷物の種類が不明のままスクリプトを適用しようとすると、全ての種類のスクリプトをテキストデータに適用し、適正な結果を採用するなどの手順を採用しなくてはならない。しかし、全ての種類のスクリプトをテキストデータに適用すると、処理に時間がかかるてしまう。

20

#### 【 0 0 2 8 】

そこで、本実施形態においては、印刷物の種類を判定した上で適用すべきスクリプトを決定する構成が採用されている。このような処理を実行するため、情報処理プログラム 1 1 は、取得部 1 1 a 、判定部 1 1 b 、適用部 1 1 c を備えている。すなわち、情報処理プログラム 1 1 が実行されると、プロセッサー 1 0 は、取得部 1 1 a 、判定部 1 1 b 、適用部 1 1 c として機能する。

#### 【 0 0 2 9 】

取得部 1 1 a により、プロセッサー 1 0 は、印刷データに基づいて印刷物を印刷するレシートプリンターから、印刷データを取得する機能を実行する。すなわち、レシートプリンター 2 から印刷データが送信されると、プロセッサー 1 0 は取得部 1 1 a の機能により、通信部 3 0 を介して印刷データを取得する。印刷データが取得されると、印刷データ 2 0 a として H D D 2 0 に記録される。

30

#### 【 0 0 3 0 】

判定部 1 1 b により、プロセッサー 1 0 は、印刷データに含まれるテキストデータに基づいて、印刷物の種類を判定する機能を実行する。本実施形態においてプロセッサー 1 0 は、テキストデータと、印刷物の種類とを対応づけた教師データに基づいて機械学習された学習モデルにより、印刷物の種類を判定する。

#### 【 0 0 3 1 】

具体的には、本実施形態における機械学習は、ニューラルネットワークによる学習である。図 2 は、本実施形態において利用されるニューラルネットワークの構造を模式的に示した図である。図 2 においては、ニューロンを模したノードを丸で示し、ノード間の結合を実線の直線で示している（ただし、図を簡易化するためノードおよび結合は一部のみが示されている）。また、図 2 においては、同一層に属するノードを縦に並べて示しており、左端の層が入力層 L i 、右端の層が出力層 L o である。

40

#### 【 0 0 3 2 】

ここでは、2 層の関係において入力層 L i に近い層を上位層、出力層 L o に近い層を下位層と呼ぶ。すなわち、ある層においては、1 層上位の層の出力を入力とし、重みの乗算とバイアスの和算を行い、活性化関数を介して出力をを行う構造となっている。例えば、図 2 に示す層 L P のノードの数が P 個であり、層 L P の各ノードに対する入力データが  $\times 1$  , ,

50

$\times M$ であり、バイアスが  $b_1, \dots, b_P$  であり、層  $L_P$  の  $k$  番目のノードに対する重みが  $w_{k1}, \dots, w_{kM}$ 、活性化関数が  $h(u)$  である場合、当該層  $L_P$  の各ノードにおける中間出力  $u_1, \dots, u_P$  は以下の式 1 で表現される。

【数 1】

$$(u_1 \cdots u_k \cdots u_P) = (x_1 \cdots x_k \cdots x_M) \begin{pmatrix} w_{11} & \cdots & w_{k1} & \cdots & w_{P1} \\ \vdots & & \vdots & & \vdots \\ w_{1k} & \cdots & w_{kk} & \cdots & w_{Pk} \\ \vdots & & \vdots & & \vdots \\ w_{1M} & \cdots & w_{kM} & \cdots & w_{PM} \end{pmatrix} + (b_1 \cdots b_k \cdots b_P) \quad \dots\dots (1)$$

10

従って、式 1 を活性化関数  $h(u_k)$  に代入すると層  $L_P$  の各ノードにおける出力  $y_1, \dots, y_P$  が得られる。

【0033】

モデルは、このように一般化されるノードが複数層設けられることによって構成される。従って、本実施形態において、モデルは、少なくとも各層における重みとバイアスと活性化関数を示すデータによって構成される。なお、活性化関数は、種々の関数、例えば、シグモイド関数や、双曲線関数 ( $tanh$ )、ReLU (Rectified Linear Unit) や他の関数を使用可能であり、非線形関数であることが好ましい。もちろん、機械学習を行う際に必要な他の条件、例えば、最適化アルゴリズムの種類や学習率等のパラメーターがモデルに含まれていても良い。また、層の数やノードの数、相関の結合関係は種々の態様とすることができる。

20

【0034】

本実施形態においては、印刷データ 20a に含まれるテキストデータを所定の入力値に変換して入力層  $L_i$  に入力し、出力層  $L_o$  から印刷物の種類を示す出力値を出力するモデルが構築される。本実施形態においては、テキストデータにキーワードが含まれているか否かを示すフラグを入力層  $L_i$  への入力値とし、印刷物が特定の種類である可能性を示す値を出力層  $L_o$  の出力値とするモデルが構築される。例えば、図 2 においては、テキストデータに "cash" というキーワードが存在する場合にノード  $I_{i2}$  に数値 1 が入力され、"cash" というキーワードが存在しない場合にノード  $I_{i2}$  に数値 0 が入力される。また、テキストデータに "grand total" というキーワードが存在する場合にノード  $I_{iM}$  に数値 1 が入力され、"grand total" というキーワードが存在しない場合にノード  $I_{iM}$  に数値 0 が入力される。すなわち、図 2 に示す入力層  $L_i$  の各ノードには、各ノードに対応するキーワードが併記されている。

30

【0035】

キーワードの有無を判定するための手法は、種々の手法であってよく、本実施形態においては、特定のフォーマットでの表記された数値もキーワードとなり得る。例えば、図 2 に示す "\$0.0\$" は、ドル通貨を示す記号の右側に、任意桁数の任意の数値、コンマ、任意桁数の任意の数値が並ぶ文字列を示しており、このフォーマットでの数値が存在する場合にキーワードが存在するとされる。

40

【0036】

出力層  $L_o$  のノード  $T_{i1} \sim T_{i3}$  は、本実施形態において印刷物の種類が 3 種類であることに応じて設けられた 3 個のノードであり、ノード  $T_{i1}, T_{i2}, T_{i3}$  のそれぞれが、購入記録を示す印刷物、売上集計を示す印刷物、勤怠情報を示す印刷物に対応している。すなわち、学習モデルの学習が完了した後においては、テキストデータが購入記録を示す印刷物のデータである場合、ノード  $T_{i1}$  の出力値が他のノードの出力値より大きくなる。テキストデータが売上集計を示す印刷物、勤怠情報を示す印刷物それぞれのデータである場合、ノード  $T_{i2}, T_{i3}$  の出力値のそれぞれが他のノードの出力値より大きくなる。なお、本実施形態においては、ノード  $T_{i1}, T_{i2}, T_{i3}$  の出力値の和が 1 になるように規格化されて

50

いる。従って、各ノードの出力の最大値は1である。

#### 【0037】

以上のモデルによれば、テキストデータに含まれるキーワードの有無を示す入力値をモデルに入力し、各ノードの出力値を計算することで、出力値が大きい順にノードに対応した種類の印刷物である可能性が高いと推定することができる。

#### 【0038】

学習済のモデルは学習モデル20bとしてHDD20に記録されるが、印刷データに含まれるテキストデータに基づいて印刷物の種類を推定する前に、当該学習モデル20bが学習されている必要がある。当該学習は、教師データ20cに基づいて行われる。すなわち、教師データ20cは、テキストデータと、当該テキストデータによって印刷される印刷物の前記種類とを対応づけたデータであり、学習するに充分な量のデータが予め収集される。

10

#### 【0039】

プロセッサー10は、判定部11bの機能により、任意のタイミングで教師データ20cに基づいて学習を実行する。この際、プロセッサー10は、図2に示すモデルによって、教師データ20cによって対応づけられたテキストデータと印刷物の種類との関係が再現されるようにモデルを調整する。すなわち、プロセッサー10は、教師データ20cに含まれる少なくとも一部のテキストデータのそれぞれに基づいて、キーワードの有無を特定し、各テキストデータを入力値に変換する。

20

#### 【0040】

さらに、プロセッサー10は、当該入力値をモデルに入力し、得られた出力値が示す印刷物の種類と、教師データ20cが示す印刷物の種類とが一致するように（差分が最小化されるように）モデルを調整する。当該調整によって、予め決められた学習完了判定基準を満たした場合に、得られたモデルが学習モデル20bとしてHDD20に記録される。

#### 【0041】

プロセッサー10は、学習モデル20bがHDD20に記録された状態で印刷物の種類の判定を行う。すなわち、プロセッサー10は、印刷データ20aを取得すると、判定部11bの機能により印刷データ20aからテキストデータを抽出する。また、プロセッサー10は、当該テキストデータからキーワードを検索し、各キーワードについての有無を特定する。さらに、プロセッサー10は、キーワードの有無をフラグ（1または0）で示す入力値を生成し、学習モデル20bに基づいて入力値に対応する出力値を取得する。そして、プロセッサー10は各ノードの出力値を比較し、出力層の値が大きい順にノードに対応する印刷物の種類である可能性が高いと見なす。

30

#### 【0042】

このようにして印刷データで印刷される印刷物の種類が上述の3種類のそれぞれである可能性が順位付けされることで、印刷物の種類が判定される。印刷物の種類が判定されると、適用部11cの機能によりプロセッサー10は、複数のスクリプトから、印刷物の種類に対応したスクリプトを選択し、選択したスクリプトをテキストデータに対して適用する。

#### 【0043】

本実施形態においては、印刷データで印刷される印刷物の種類である可能性が高い順が特定されているため、プロセッサー10は、可能性が高い種類であるほど順位が高くなるように、各印刷物の種類に対応したスクリプトを選択する優先順位を設定する。そして、プロセッサー10は、選択したスクリプトをテキストデータに対して適用して結果が得られるまで、設定した優先順位に従って、スクリプトを選択して適用する。

40

#### 【0044】

すなわち、本実施形態におけるスクリプトは、特定のフォーマットのテキストを検出したり、特定のキーワードと項目名との組み合わせを検出したりして、特定のパラメーターに対してパラメーターに対応する項目を対応づける処理の手順である。プロセッサー10は、スクリプトが適用された場合に得られた結果が既定の判断基準を満たす場合にスクリプトの適用結果が得られたと判定する。既定の判断基準は、種々の手法で決められてよいが

50

、本実施形態においては、スクリプトの適用によって得られた構造化データにおいて、予め必須とされた項目のパラメーターが得られたか否かが判断基準となっている。

#### 【 0 0 4 5 】

すなわち、必須とされた項目のパラメーターが構造化データとして取得された場合に、プロセッサー 10 は、スクリプトの結果が得られたと判定する。必須とされた項目のパラメーターが構造化データとして取得されなかった場合、プロセッサー 10 は、スクリプトの結果が得られなかったと判定する。むろん、必須とされる項目は複数個であってもよい。

#### 【 0 0 4 6 】

プロセッサー 10 は、このようなスクリプトの適用と結果が得られたか否かの判定を、優先順位に従って実施する。すなわち、プロセッサー 10 は、スクリプトデータ 20 d を参照し、最も優先順位が高いとされた印刷物の種類に対応するスクリプトを印刷データ 20 a のテキストデータに適用し、構造化データを取得する。プロセッサー 10 は、当該構造化データに基づいてスクリプトの結果が得られたか否かを判定し、結果が得られた場合、プロセッサー 10 は、スクリプトの適用を終了する。

10

#### 【 0 0 4 7 】

一方、結果が得られなかった場合、プロセッサー 10 は、次の優先順位のスクリプトを印刷データ 20 a のテキストデータに適用し、構造化データに基づいて結果が得られたか否か判定する。プロセッサー 10 は、以後、結果が得られたと判定されるまで、順次、優先順位を 1 個下げてスクリプトを適用し、結果が得られたか否か判定する処理を繰り返す。

20

#### 【 0 0 4 8 】

以上のような本実施形態によれば、プロセッサー 10 は、印刷データ 20 a に含まれるテキストデータに基づいて印刷物の種類に対応したスクリプトを選択し、適用する。また、本実施形態においては、テキストデータで印刷される印刷物の種類である可能性が高い順にスクリプトを適用していく。この結果、全ての種類のスクリプトを適用しなくても、適切なスクリプトをテキストデータに対して適用できる可能性を高めることができる。

30

#### 【 0 0 4 9 】

なお、以上の処理によって印刷データ 20 a から構造化データが得られると、当該データは構造化データ 20 e として HDD 20 に記録され、構造化データ 20 e は各種の解析のために利用可能になる。例えば、複数の印刷データ 20 a の構造化データ 20 e が I S V ( independent software vendor ) 等に提供され、各種の解析やソフトウェアの開発等に利用される。

30

#### 【 0 0 5 0 】

##### ( 2 ) 機械学習処理 :

次に、プロセッサー 10 が実行する機械学習処理を説明する。図 3 は機械学習処理を示すフローチャートである。機械学習処理は、印刷物の種類を判定する前に予め実行される。機械学習処理が開始されると、プロセッサー 10 は、判定部 11 b の機能により、訓練モデルを取得する ( ステップ S 100 ) 。すなわち、予め決められたニューラルネットワークのパラメーター ( 重みやバイアス等 ) を訓練モデルとして取得する。

40

#### 【 0 0 5 1 】

訓練モデルにおいては、テキストデータにおけるキーワードの有無を入力値として、印刷物の種類を示す出力値を出力するが、出力値が示す印刷物の種類は初期において正確ではない。すなわち、訓練モデルにおいては、ノードが構成する層の数やノードの数は決められるが、入出力の関係を規定するパラメーター ( 上述の重みやバイアス等 ) は最適化されていない。これらのパラメーターは、機械学習の過程で最適化される ( すなわち、訓練される ) 。本実施形態において訓練モデルは予め決定されているが、利用者がマウス 3 a やキーボード 3 b を操作して決定しても良い。

#### 【 0 0 5 2 】

次に、プロセッサー 10 は、判定部 11 b の機能により、教師データ 20 c を取得する ( ステップ S 105 ) 。すなわち、プロセッサー 10 は、予め HDD 20 に記録されている教師データ 20 c を取得する。なお、教師データ 20 c は、レシートプリンター 2 の運用

50

過程や、新たなフォーマットの印刷物の使用が開始される前等において適宜追加されてよい。もちろん、既存のフォーマットの印刷物の廃止等に応じて廃止される印刷物の教師データ<sub>20c</sub>が削除されてもよい。

#### 【0053】

次に、プロセッサー10は、判定部11bの機能により、テストデータを取得する（ステップS110）。本実施形態においては、教師データ<sub>20c</sub>の一部がテストデータとされる。テストデータは、教師データ<sub>20c</sub>と区別される。なお、テストデータの情報量と教師データ<sub>20c</sub>の情報量は、種々の量であって良いが、本実施形態においては、教師データ<sub>20c</sub>の方がテストデータより多くなるように設定される。

#### 【0054】

次に、プロセッサー10は、判定部11bの機能により、初期値を決定する（ステップS115）。すなわち、プロセッサー10は、ステップS100で取得した訓練モデルのうち、可変のパラメーターに対して初期値を与える。初期値は、種々の手法で決定されて良い。例えば、ランダム値や0等を初期値とすることができます、重みとバイアスとで異なる思想で初期値が決定されても良い。もちろん、学習の過程でパラメーターが最適化されるように初期値が調整されても良い。

10

#### 【0055】

次に、プロセッサー10は、判定部11bの機能により、学習を行う（ステップS120）。すなわち、プロセッサー10は、教師データ<sub>20c</sub>の各テキストデータを抽出し、各テキストデータにおけるキーワードの有無に基づいて各テキストデータを入力値に変換する。そして、プロセッサー10は、ステップS100で取得した訓練モデルに当該入力値を入力し、印刷物の種類毎に、各種類である可能性を示す出力値を計算する。また、プロセッサー10は、出力値と、各テキストデータに対応づけられた印刷物の種類を示すフラグと、の差を示す損失関数によって誤差を特定する。そして、プロセッサー10は、損失関数のパラメーターによる微分に基づいてパラメーターを更新する処理を既定回数繰り返す。

20

#### 【0056】

教師データ<sub>20c</sub>に含まれるサンプルの数を*i*で表現した場合、プロセッサー10は、*i*番目のサンプルのテキストデータから特定された入力値を訓練モデルに入力することで、*i*番目のサンプルの出力値（T<sub>i1</sub>～T<sub>i3</sub>）を取得する。一方、教師データ<sub>20c</sub>に対応づけられた印刷物の種類を示す情報は、印刷物の種類に対応した値（t<sub>i1</sub>～t<sub>i3</sub>）であり、特定の印刷物の種類に対応した値が1、他の種類に対応した値が0となる。

30

#### 【0057】

そこで、訓練モデルの出力値をT<sub>i</sub>と表記し、教師データ<sub>20c</sub>が示す印刷物の種類をt<sub>i</sub>と表記すると、*i*番目のサンプルについての損失関数はL（T<sub>i</sub>, t<sub>i</sub>）と表記可能である。もちろん、損失関数は、種々の関数を採用可能であり、例えば、交差エントロピー誤差などを採用可能である。以上のような損失関数Lを算出する処理は、教師データ<sub>20c</sub>が示すサンプルの全てまたは一部について実施され、その平均や総和によって1回の学習における損失関数が表現される。例えば、総和によって損失関数が表現される場合、全体の損失関数Eは次の式2で表現される。

40

#### 【数2】

$$E = \sum_i L_i(T_i, t_i) \quad \dots \quad (2)$$

損失関数Eが得られたら、プロセッサー10は、既定の最適化アルゴリズム、例えば、確率的勾配降下法等によってパラメーターを更新する。

#### 【0058】

以上のようにして、既定回数のパラメーターの更新が行われると、プロセッサー10は、訓練モデルの汎化が完了したか否かを判定する（ステップS125）。すなわち、プロセッサー10は、ステップS110で取得したテストデータから算出される入力値を訓練モ

50

デルに入力して出力値を取得する。そして、プロセッサー 10 は、最も大きい出力値に対応する印刷物の種類と、テストデータに対応づけられた印刷物の種類とが一致している数を取得し、テストデータのサンプル数で除することで推定精度を取得する。本実施形態において、プロセッサー 10 は、推定精度が閾値以上である場合に汎化が完了したと判定する。

#### 【 0 0 5 9 】

なお、汎化性能の評価に加え、ハイパーパラメーターの妥当性の検証が行われてもよい。すなわち、重みとバイアス以外の可変量であるハイパーパラメーター、例えば、ノードの数等がチューニングされる構成において、プロセッサー 10 は、検証データに基づいてハイパーパラメーターの妥当性を検証しても良い。検証データは、例えば、教師データ 20c から予め抽出され、教師データ 20c やテストデータと区別されることによって用意される。

10

#### 【 0 0 6 0 】

ステップ S 125において、訓練モデルの汎化が完了したと判定されない場合、プロセッサー 10 は、ステップ S 120 を繰り返す。すなわち、さらに重みおよびバイアスを更新する。一方、ステップ S 125において、訓練モデルの汎化が完了したと判定された場合、プロセッサー 10 は、学習モデルを記録する（ステップ S 130）。すなわち、プロセッサー 10 は、訓練モデルを学習モデル 20b として HDD 20 に記録する。以上のようにして得られた学習モデル 20b で印刷物の種類を推定する構成によれば、任意の印刷データに含まれるテキストデータに基づいて、印刷データで印刷される印刷物の種類を推定することが可能になる。

20

#### 【 0 0 6 1 】

##### （ 3 ）情報処理 :

次に、プロセッサー 10 が実行する情報処理を説明する。図 4 は、情報処理を示すフローチャートである。本実施形態において情報処理は、任意のレシートプリンター 2 から印刷データが送信された場合に実行される。

#### 【 0 0 6 2 】

情報処理が開始されると、プロセッサー 10 は、取得部 11a の機能により、印刷データを取得する（ステップ S 200）。すなわち、プロセッサー 10 は、取得部 11a の機能により、通信部 30 を介してレシートプリンター 2 と通信を行い、レシートプリンター 2 が送信した印刷データを受信する。プロセッサー 10 は、受信した印刷データを印刷データ 20a として HDD 20 に記録する。

30

#### 【 0 0 6 3 】

次に、プロセッサー 10 は、判定部 11b の機能によりキーワード検索を行う（ステップ S 205）。すなわち、プロセッサー 10 は、ステップ S 200 で受信された印刷データ 20a からテキストデータを抽出し、予め決められたキーワードが存在するか否か判定する。プロセッサー 10 は、存在したキーワードに 1、存在しなかったキーワードに 0 を対応づける。例えば、図 2 に示す例であれば、"\$ 0 . 0" で示されたフォーマットのキーワードが存在する場合にフラグを 1、"c a s h" というキーワードが存在する場合にフラグを 1、"h o u r s" というキーワードが存在しない場合にフラグを 0 に設定するなどの処理を行う。

40

#### 【 0 0 6 4 】

次に、プロセッサー 10 は、学習モデルによる推定を行う（ステップ S 210）。すなわち、プロセッサー 10 は、学習モデル 20b に基づいて、ステップ S 205 で設定されたフラグを入力値としたニューラルネットワークの演算を行う。この結果、印刷物の種類に応じた各ノードについての出力値が得られる。例えば、図 2 に示す例において、ステップ S 205 で設定されたフラグが入力値とされ、学習モデル 20b による演算が行われると、ノード T<sub>i1</sub>, T<sub>i2</sub>, T<sub>i3</sub> のそれぞれについて 0 以上 1 以下の出力値が得られる。

#### 【 0 0 6 5 】

次に、プロセッサー 10 は、優先順位を取得する（ステップ S 215）。すなわち、プロ

50

セッサー 10 は、各ノードの出力値の大きさが、各ノードに対応する種類の印刷物である可能性を示していると見なす。そして、プロセッサー 10 は、出力値が大きい順に各ノードに対応する種類を並べ、当該種類に対応したスクリプトを適用する際の優先順位とする。

#### 【 0 0 6 6 】

例えば、図 2 に示す例においては、ノード  $T_{i1}$ ,  $T_{i2}$ ,  $T_{i3}$  のそれぞれが、購入記録を示す印刷物、売上集計を示す印刷物、勤怠情報を示す印刷物に対応している。この例において、ノード  $T_{i1}$ ,  $T_{i2}$ ,  $T_{i3}$  の出力値が 0.1, 0.85, 0.05 である場合、売上集計を示す印刷物に対応したスクリプトの優先順位が最も高く、次いで、購入記録を示す印刷物に対応したスクリプト、次に、勤怠情報を示す印刷物に対応したスクリプトという優先順位になる。

#### 【 0 0 6 7 】

次に、プロセッサー 10 は、未適用で最も優先順位が高いスクリプトを適用し（ステップ S 220）、結果が得られたと判定されるまでステップ S 220 を繰り返す（ステップ S 225）。すなわち、プロセッサー 10 は、ステップ S 200 で取得された印刷データからテキストデータを抽出し、ステップ S 215 で取得された優先順位が高い順にスクリプトをテキストデータに適用する。スクリプトがテキストデータに対して適用されると、構造化データが得られるため、プロセッサー 10 は、当該構造化データに基づいて、既定の判断基準を満たすか否かを判定する。そして、プロセッサー 10 は、既定の判断基準を満たす場合に、結果が得られたと判定する。ステップ S 225において、結果が得られたと判定された場合、プロセッサー 10 は、得られた構造化データ 20e を HDD 20 に保存する（ステップ S 230）。以上の構成によれば、テキストデータで印刷される印刷物の種類である可能性が高い順にスクリプトを適用していく。この結果、全ての種類のスクリプトを適用しなくとも、適切なスクリプトをテキストデータに対して適用できる可能性を高めることができる。

#### 【 0 0 6 8 】

##### （ 4 ）他の実施形態：

以上の実施形態は本発明を実施するための一例であり、他にも種々の実施形態を採用可能である。例えば、本発明の一実施形態にかかる情報処理装置は、種々の態様で提供されてよく、クラウド型のサーバーなど、複数の装置によって提供されてもよい。さらに、情報処理装置 1 によって実行される機能の少なくとも一部がレシートプリンター 2 で実行されてもよい。例えば、学習モデル 20b がレシートプリンター 2 に保存され、レシートプリンター 2 において印刷物の種類が判定されてもよい。さらに、以上の実施形態のように印刷データに含まれるテキストデータに基づいて、印刷データで印刷される印刷物の種類を判定する手法は、プログラムの発明、方法の発明としても実現可能である。

#### 【 0 0 6 9 】

さらに、機械学習と、印刷物の種類の判定とが別の装置で実行されてもよい。さらに、上述の実施形態は一例であり、一部の構成が省略されたり、他の構成が追加されたりする実施形態が採用され得る。

#### 【 0 0 7 0 】

取得部は、印刷データに基づいて印刷物を印刷するレシートプリンターから、印刷データを取得することができればよい。すなわち、取得部は、任意の印刷物を印刷させるための印刷データを、任意のタイミングで取得することができればよい。レシートプリンターと情報処理装置との接続態様は種々の態様であってよく、ケーブルによって直接的に接続されていてもよいし、インターネット等を介して接続可能であってもよい。

#### 【 0 0 7 1 】

レシートプリンターは、レシートを印刷することができればよく、むろん、レシート以外の印刷物が印刷可能であって良い。レシートは、店舗での購入記録を示す印刷物であれば良く、購入記録が印刷される限り、印刷内容やレイアウトは限定されない。印刷データは、レシートプリンターに任意の印刷物を印刷させるためのデータであれば良く、任意のフォーマットであって良い。ただし、印刷データにはテキストデータが含まれ、当該テキス

トデータが示すテキストが印刷用紙に印刷されて印刷物となる。むろん、印刷物にはテキスト以外のオブジェクトが印刷されてよく、各種の画像や墨線等を印刷させるためのデータが印刷データに含まれていてもよい。

#### 【 0 0 7 2 】

なお、テキストデータは、印刷物に印刷される文字列を示していればよいが、むろん、テキストのレイアウトやフォントの種類、大きさ等の情報が含まれていてもよい。印刷物の種類の判定にこれらの情報が利用されても良い。

#### 【 0 0 7 3 】

判定部は、印刷データに含まれるテキストデータに基づいて、印刷物の種類を判定することができればよい。すなわち、テキストデータは、印刷物に印刷されるテキストを示しており、テキストによって印刷物の内容が記述される。従って、テキストを印刷物の種類を判定するための要素として利用することができる。印刷物の種類を判定するための手法は、上述の実施形態のように機械学習による学習済のモデルを利用する構成に限定されない。例えば、テキストデータが示すテキスト内に印刷物の種類に応じた特定のキーワードが含まれている比率等によって印刷物の種類が判定される構成等が採用されてもよい。

10

#### 【 0 0 7 4 】

印刷物の種類は、印刷物の用途等によって予め分類されていればよく、上述の実施形態のように、店舗での購入記録を示す印刷物、店舗の売上集計を示す印刷物、勤怠情報を示す印刷物以外にも種々の印刷物が挙げられる。例えば、購入された宝くじやチケット等を示す印刷物であってもよいし、公共料金の支払い証明を示す印刷物であってもよく、他にも種々の印刷物であってよい。

20

#### 【 0 0 7 5 】

適用部は、複数のスクリプトから、印刷物の種類に対応したスクリプトを選択し、選択したスクリプトをテキストデータに対して適用することができればよい。すなわち、印刷物の種類は複数個存在し、それぞれに応じたスクリプトが予め用意される。スクリプトは、印刷物の種類に応じて用意されるため、少なくとも印刷物の種類の数と同一数だけ存在するが、1種類の印刷物に対して複数のスクリプトが予め用意されてもよい。

#### 【 0 0 7 6 】

スクリプトは、テキストデータに基づいて既定の手順で処理を行うためのプログラムであれば良く、言語は限定されない。スクリプトは簡易な処理の手順であることが多いが、むろん、処理の内容は複雑であってもよい。スクリプトを適用する目的は、種々の目的であってよく、スクリプト毎に目的が存在してもよい。目的としては、上述の実施形態のように、項目とパラメーターとを対応づけて抽出し、各種の解析の元データとして提供する目的以外にも、種々の目的が想定される。例えば、購入記録に含まれる商品毎の金額や合計金額等に基づいて家計管理のための情報を収集する目的や、店舗の売上集計に基づいて市場調査のための情報を収集する目的や、勤怠情報に基づいて労務管理のための情報を収集する目的など、各種の目的が想定される。

30

#### 【 0 0 7 7 】

スクリプトのテキストデータに対する適用は、スクリプトが示す手順に従ってテキストデータを解析する処理を実行することに相当する。ここでは、テキストデータに基づいて既定の処理が実行されればよく、スクリプトはインターフリタ型であってもよいしコンパイル型であってもよい。

40

#### 【 0 0 7 8 】

教師データは、印刷データに含まれるテキストデータと、当該印刷データによって印刷される印刷物の種類とを対応づけていればよい。教師データは、学習モデルの入力値および出力値のフォーマットと同一のフォーマットで定義されなければよく、上述の実施形態のようなフォーマットに限定されない。例えば、テキストデータが示す文字列が入力値とされてもよいし、テキストデータに含まれ得る文字列を示すone-hotベクトルが定義されて入力値とされてもよいし、テキストデータに含まれる単語や単語の数が入力値とされてもよく、種々の構成を採用可能である。

50

**【 0 0 7 9 】**

テキストデータから導出される値が入力値となる場合、当該値の導出は種々の手法で行われてよい。例えば、文字列や単語が検索され、検索結果の有無や検索された数が取得されてもよい。さらに、通貨や時刻が特定のフォーマットで表現されているか否か判定されてよい。例えば、文字列の左側に通貨を示す記号が存在し、その右側に数桁の数値が存在し、その右側にコンマが存在し、さらにその右側に数値が存在するような文字列を検索して価格の有無が特定されてもよい。むろん、このような検索に際しては、正規表現が利用されて良い。

**【 0 0 8 0 】**

また、出力値は、印刷物の複数の種類について、それぞれの種類である可能性を示していくてもよいし、それぞれの種類ではない可能性を示していくてもよい。また、印刷物の種類毎に学習モデルが生成され、印刷物の種類毎に当該種類である可能性を示す値が出力値とされてもよく、他にも種々の態様が想定可能である。むろん、出力値がスクリプトを示していくてもよい。

10

**【 0 0 8 1 】**

機械学習の手法は、種々の手法であって良い。すなわち、テキストデータ自体やテキストデータから導出される値（単語の数やテキストの長さ、スペースの数、人名の有無等）を入力し、印刷物の種類を出力するモデルを構築し、当該モデルによって適正な出力が得られるように機械学習すればよい。機械学習においては、当該モデルによる出力と教師データにおける出力値との差分を極小化する学習を行うことができればよい。

20

**【 0 0 8 2 】**

従って、例えばニューラルネットワークによる機械学習が行われる場合、モデルを構成する層の数やノードの数、活性化関数の種類、損失関数の種類、勾配降下法の種類、勾配降下法の最適化アルゴリズムの種類、ミニバッチ学習の有無やバッチの数、学習率、初期値、過学習抑制手法の種類や有無、畳み込み層の有無、畳み込み演算におけるフィルタのサイズ、フィルタの種類、パディングやストライドの種類、プーリング層の種類や有無、全結合層の有無、再帰的な構造の有無など、種々の要素を適宜選択して機械学習が行われればよい。むろん、他の機械学習、例えば、サポートベクターマシンやクラスタリング、強化学習等によって学習が行われてもよい。

30

**【 0 0 8 3 】**

さらに、モデルの構造（例えば、層の数や層毎のノードの数等）が自動的に最適化される機械学習が行われてもよい。さらに、情報処理装置において機械学習が行われる構成において、複数のレシートプリンターから教師データが収集され、この教師データに基づいて機械学習が行われる構成であっても良い。

**【 0 0 8 4 】**

むろん、テキストデータがRNN (Recurrent Neural Network) によって処理されても良い。すなわち、RNNによれば、テキストデータの前後関係を考慮することが可能であるため、RNNによる処理が含まれるモデルの機械学習が利用されても良い。図6は、RNNを含むモデルの構成例である。図6においては、入力値となるテキストデータが単語に分解され、入力値とされる。図6においては、購入記録を構成するテキストの一部である"Cashier Name A Milk \$ 4.99, , , 0.70"という連続した文字列が単語に分解され、"Cashier"が入力値×1、"Name A"が入力値×2、"Milk"が入力値×3、"\$"が入力値×4、"4.99"が入力値×5とされ、以後、"0.70"が入力値×nと変換される例が示されている。

40

**【 0 0 8 5 】**

さらに、図6においては、各入力値がRNNのブロックM<sub>1</sub>～M<sub>5</sub>に入力されて出力値y<sub>1</sub>～y<sub>5</sub>、～y<sub>n</sub>が出力される例が想定されている。なお、図6に示す例においては、RNNのブロックM<sub>1</sub>～M<sub>5</sub>、～M<sub>n-1</sub>の出力値y<sub>1</sub>～y<sub>5</sub>、～y<sub>n-1</sub>が次のブロックM<sub>2</sub>～M<sub>6</sub>、～M<sub>n</sub>に入力される。なお、最初のRNNのブロックM<sub>1</sub>に対して入力値x<sub>1</sub>とともに入力される値y<sub>0</sub>は、種々の手法で初期化されて良い（例えば、0ベクトル等）。図6

50

に示す例においては、これらの出力値がニューラルネットワーク（NN）に入力され、最終的に印刷物の種類に対応したノードT<sub>i1</sub>，T<sub>i2</sub>，T<sub>i3</sub>毎の出力値が得られる。

#### 【0086】

以上のようなモデルにおいて、入力値の定義法や重み、バイアスなどのRNNのモデルやNNのモデルは、公知の種々の手法を用いることができる。図1に示す例において、以上のようなモデルが想定され、プロセッサー10が判定部11bの機能によって機械学習を行えば、テキストデータを入力し、印刷データによって印刷される印刷物の種類を推定するモデルを機械学習することができる。そして、このような構成においても、印刷物の種類を判定し、判定結果に応じたスクリプトを適用することにより、全ての種類のスクリプトを適用しなくとも、適切なスクリプトをテキストデータに対して適用できる可能性を高めることができる。もちろん、図6に示すRNNのモデルも一例であり、双方向RNNが採用されるなど、種々のモデルが採用されてよい。

10

#### 【0087】

印刷物が複数の種類のそれである可能性は、種々の手法によって定義されてよい。従って、印刷物が特定の種類である可能性の最大値が1であり、最小値が0である上述の実施形態以外にも種々の構成が採用されてよい。いずれにしても、印刷物が特定の種類であることを示す数値の大小関係によって、印刷データで印刷される印刷物が複数の印刷物の中のそれである可能性を順位付けすることができればよい。優先順位は、当該順位付けによって特定されればよい。

20

#### 【0088】

選択したスクリプトをテキストデータに対して適用して結果が得られたか否かは、予め決められた判定基準によって判断されればよい。従って、上述の実施形態のようにスクリプトが既定の構造化データを出力する構成において、構造化データの中で予め必須とされたデータが得られた場合に結果が得られたと見なす構成以外にも種々の構成が採用されてよい。例えば、スクリプトが適用された場合に、エラーの発生比率や項目とパラメーターとの対応関係が不定となった比率等が閾値以下である場合等にスクリプトの適用結果が得られたと判定する構成等が挙げられる。

#### 【0089】

さらに本発明は、コンピューターが実行するプログラムや、方法としても適用可能である。また、以上のようなプログラム、方法は、単独の装置として実現される場合もあれば、複数の装置が備える部品を利用して実現される場合もあり、各種の態様を含むものである。また、一部がソフトウェアであり一部がハードウェアであったりするなど、適宜、変更可能である。さらに、プログラムの記録媒体としても発明は成立する。もちろん、そのプログラムの記録媒体は、磁気記録媒体であってもよいし半導体メモリ等であってもよいし、今後開発されるいかなる記録媒体においても全く同様に考えることができる。

30

#### 【符号の説明】

#### 【0090】

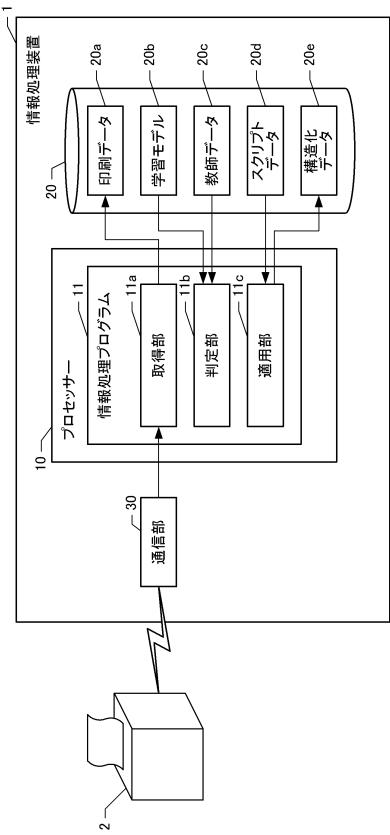
1…情報処理装置、2…レシートプリンター、3a…マウス、3b…キーボード、10…プロセッサー、11…情報処理プログラム、11a…取得部、11b…判定部、11c…適用部、20…HDD、20a…印刷データ、20b…学習モデル、20c…教師データ、20d…スクリプトデータ、20e…構造化データ、30…通信部

40

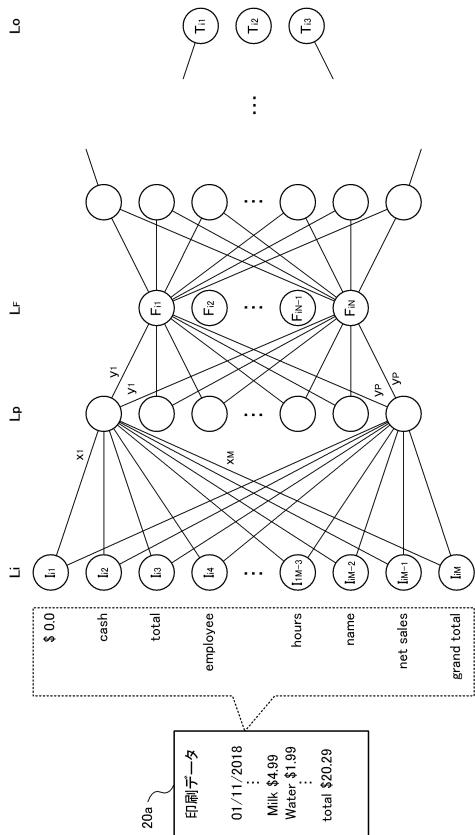
50

【図面】

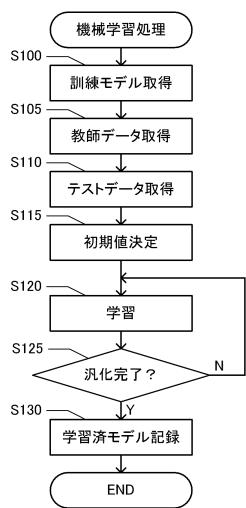
【図 1】



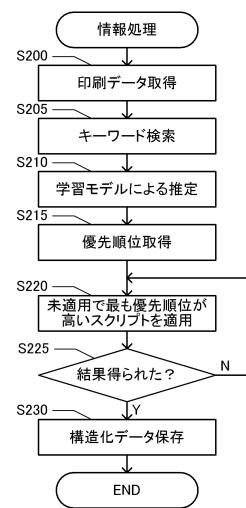
【図 2】



【図 3】



【図 4】



10

20

30

40

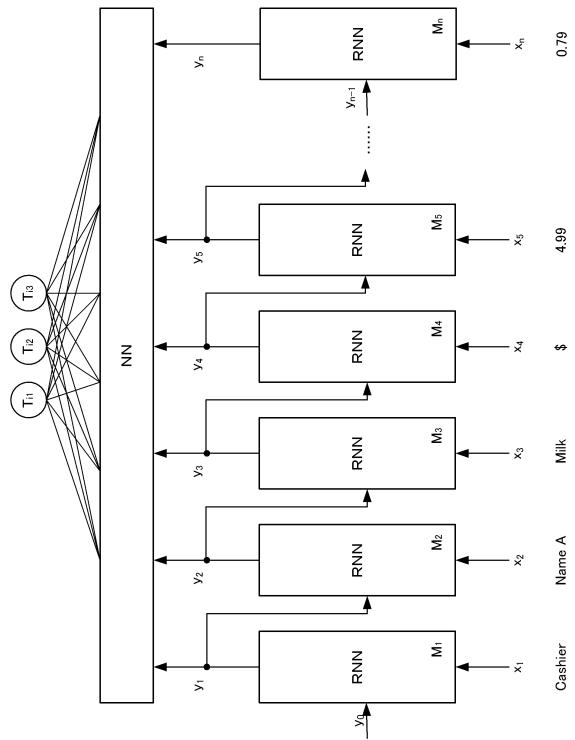
50

【図 5】

↙ Pr ↘ St

Retail Receipt		
Chk# 334589 02/17/2017		
Cashier: Name A		
Milk	2% 1 gallon	\$ 4.99
Cola	1 liter bottle	\$ 1.29
Blueberries, 2.00 lb	Organic, \$21.00/lb	\$ 42.00
Coffee	Colombia, medium roast, 7oz	\$ 7.39
Milk	2% 1 gallon	\$ 4.99
Apples, 1.00 lb	Gala, \$1.50/lb	\$ 1.50
SUBTOTAL		\$ 62.16
13% HST		\$ 8.05
TOTAL		\$70.21
Cash tendered: \$ 80.00		
Change:		
		\$ 0.79

【図 6】



---

フロントページの続き

(51)国際特許分類

F I

B 4 1 J 29/38  
G 0 6 N 20/00 1 3 0

(56)参考文献 特開2018-018465(JP,A)

(58)調査した分野 (Int.Cl., DB名)

G 0 6 F 3 / 0 9 - 3 / 1 2  
H 0 4 N 1 / 0 0  
B 4 1 J 2 9 / 0 0 - 2 9 / 7 0  
B 4 1 J 5 / 0 0 - 5 / 5 2 ; 2 1 / 0 0 - 2 1 / 1 8  
G 0 3 G 1 3 / 3 4 ; 1 5 / 0 0 ; 1 5 / 3 6 ; 2 1 / 0 0 ; 2 1 / 0 2 ; 2 1 / 1 4 ;  
2 1 / 2 0  
G 0 6 F 1 6 / 0 0 - 1 6 / 9 5 8  
G 0 6 N 3 / 0 0 - 3 / 1 2 ; 7 / 0 8 - 9 9 / 0 0  
G 0 7 G 1 / 0 0 - 5 / 0 0