



(12)发明专利申请

(10)申请公布号 CN 111210218 A

(43)申请公布日 2020.05.29

(21)申请号 201911134738.6

(22)申请日 2019.11.19

(30)优先权数据

2018-218996 2018.11.22 JP

(71)申请人 精工爱普生株式会社

地址 日本东京

(72)发明人 大宫英久

(74)专利代理机构 北京金信知识产权代理有限公司

公司 11225

代理人 敖敦格日乐 权太白

(51)Int.Cl.

G06Q 20/20(2012.01)

G07G 1/12(2006.01)

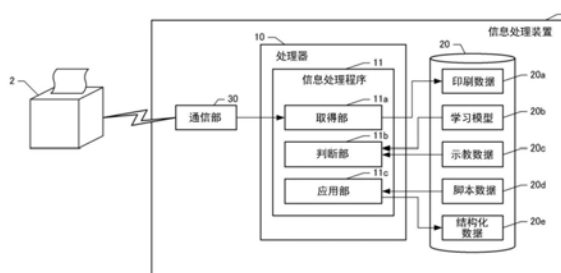
权利要求书2页 说明书12页 附图5页

(54)发明名称

信息处理装置、票据打印机以及信息处理方法

(57)摘要

本发明提供一种信息处理装置、票据打印机以及信息处理方法。该信息处理装置具备：取得部，其从基于印刷数据而对印刷物进行印刷的票据打印机，取得所述印刷数据；判断部，其基于被包含在所述印刷数据中的文本数据，而对所述印刷物的种类进行判断；应用部，其从多个脚本中选择与所述种类对应的所述脚本，且将所选择的所述脚本应用于所述文本数据。



1. 一种信息处理装置,其具备处理器,
所述处理器
从基于印刷数据而对印刷物进行印刷的票据打印机,取得所述印刷数据,
并基于被包含在所述印刷数据中的文本数据,而对所述印刷物的种类进行判断,
且从多个脚本中选择与所述种类对应的所述脚本,并且将所选择的所述脚本应用于所述文本数据。
2. 如权利要求1所述的信息处理装置,其中,
所述处理器通过学习模型来实施所述种类的判断,其中,所述学习模型为,基于将所述文本数据与所述印刷物的所述种类建立了对应的示教数据而实施了机器学习的模型。
3. 如权利要求1或权利要求2所述的信息处理装置,其中,
所述处理器
对于多个所述种类,而分别取得作为所述印刷物的所述种类的可能性,
并且,以越是所述可能性高的所述种类则使等级越高的方式,对选择与所述种类对应的所述脚本的优先等级进行设定。
4. 如权利要求3所述的信息处理装置,其中,
所述处理器按照所设定的所述优先等级而选择并应用所述脚本,直至将所选择的所述脚本应用于所述文本数据并获得结果为止。
5. 如权利要求1所述的信息处理装置,其中,
在所述印刷物的所述种类中,包含有表示在店铺的购买记录的所述印刷物以及表示店铺的销售额汇总的所述印刷物。
6. 如权利要求1所述的信息处理装置,其中,
所述脚本为,用于实施使被包含在所述文本数据中的项目名与和所述项目名对应的参数建立对应的处理的程序。
7. 一种票据打印机,具备:
印刷头,其基于所述印刷数据而对所述印刷物进行印刷;
通信接口,其能够向权利要求1至权利要求6的任意一项所述的信息处理装置发送所述印刷数据。
8. 一种信息处理方法,包括:
取得工序,其从基于印刷数据而对印刷物进行印刷的票据打印机,取得所述印刷数据;
判断工序,其基于被包含在所述印刷数据中的文本数据,而对所述印刷物的种类进行判断;
应用工序,其从多个脚本中选择与所述种类对应的所述脚本,并将所选择的所述脚本应用于所述文本数据。
9. 如权利要求8所述的信息处理方法,其中,
通过学习模型来实施所述种类的判断,其中,所述学习模型为,基于将所述文本数据与所述印刷物的所述种类建立了对应的示教数据而实施了机器学习的模型。
10. 如权利要求8或权利要求9所述的信息处理方法,其中,
对于多个所述种类,而分别取得作为所述印刷物的所述种类的可能性,
以越是所述可能性高的所述种类则使等级越高的方式,对选择与所述种类对应的所述

脚本的优先等级进行设定。

11. 如权利要求10所述的信息处理方法, 其中,

按照所设定的所述优先等级而选择并应用所述脚本, 直至将所选择的所述脚本应用于所述文本数据并获得结果为止。

12. 如权利要求8所述的信息处理方法, 其中,

在所述印刷物的所述种类中, 包含有表示在店铺的购买记录的所述印刷物以及表示店铺的销售额汇总的所述印刷物。

13. 如权利要求8所述的信息处理方法, 其中,

所述脚本为, 用于实施使被包含在所述文本数据中的项目名与和所述项目名对应的参数建立对应的处理的程序。

信息处理装置、票据打印机以及信息处理方法

技术领域

[0001] 本发明涉及一种信息处理装置、票据打印机以及信息处理方法。

背景技术

[0002] 一直以来,已知一种如下的技术,即,对用于由在POS终端等被利用的票据打印机实施印刷的印刷数据进行解析的技术。例如,在专利文献1中,公开了一种如下的技术,即,当将印刷数据发送至控制服务器时,控制服务器通过与布局对应的方法而对印刷数据进行解析的技术。

[0003] 由于在根据印刷数据而被印刷的印刷物的布局不明确的状态下,与布局相对应的解析的方法也不明确,因此能够采用通过设想到的所有的方法而实施解析并将最恰当的结果视为解析结果的方法。如上文所述在通过所有的方法而实施解析的结构中,解析将会花费时间。

[0004] 专利文献1:日本特开2015-158775号公报

发明内容

[0005] 本发明的一个实施方式中的信息处理装置具备处理器,处理器从基于印刷数据而对印刷物进行印刷的票据打印机,取得印刷数据,并且基于被包含在印刷数据中的文本数据,而对印刷物的种类进行判断,并从多个脚本中选择与种类对应的脚本,并且将所选择的脚本应用于文本数据。

[0006] 此外,也可以为如下结构,即,处理器通过学习模型来实施种类的判断,其中,所述学习模型为,基于将文本数据与印刷物的种类建立了对应的示教数据而实施了机器学习的模型。

[0007] 另外,也可以为如下的结构,即,处理器对于多个种类,而分别取得作为所述印刷物的所述种类的可能性,并且,以越是可能性高的种类则使等级越高的方式,对选择与印刷物的种类对应的脚本的优先等级进行设定。

[0008] 另外,也可以为如下结构,即,处理器按照所设定的优先等级而选择并应用脚本,直至将所选择的脚本应用于文本数据并获得结果为止。

[0009] 另外,也可以为如下结构,即,在印刷物的种类中,包含有表示在店铺的购买记录的印刷物以及表示店铺的销售额汇总的印刷物。

[0010] 另外,所述脚本也可以为用于实施使被包含在文本数据中的项目名与和项目名对应的参数建立对应的处理的程序。

附图说明

[0011] 图1是信息处理装置的框图。

[0012] 图2是表示神经网络的结构图。

[0013] 图3是表示机器学习处理的流程图。

[0014] 图4是表示信息处理的流程图。

[0015] 图5是表示示出购买记录的印刷物和结构化数据的示例的图。

[0016] 图6是包含RNN的模型的结构图。

具体实施方式

[0017] 在此,按照下述的顺序来对本发明的实施方式进行说明。

[0018] (1) 信息处理装置的结构:

[0019] (2) 机器学习处理:

[0020] (3) 信息处理:

[0021] (4) 其他实施方式:

[0022] (1) 信息处理装置的结构:

[0023] 图1是表示本发明的实施方式所涉及的信息处理装置的框图。本实施方式所涉及的信息处理装置1为能够与票据打印机2进行通信的服务器,且能够经由互联网等网络而与票据打印机2进行通信。

[0024] 票据打印机2为被设置在店铺的收款机处的打印机。即,票据打印机2为了向在店铺中购买了商品的各个顾客交付表示购买记录的印刷物(票据)而被使用,在结账后通过操作员的操作而印刷票据。当然,票据打印机2既可以为与其他的设备进行协同动作的结构,也可以构成例如POS系统(point of sales system:销售点系统)的一部分。

[0025] 票据打印机2能够根据来自未图示的收款机等控制装置的指令而对各种的印刷物进行印刷。在本实施方式中,未图示的控制装置能够对表示在店铺中的购买记录的印刷物、表示店铺的销售额汇总的印刷物、表示在店铺工作的人员的考勤信息的印刷物的任意一个进行指定并印刷。

[0026] 即,控制装置能够通过未图示的处理器功能,而生成表示购买记录的印刷数据,其中,该购买记录表示顾客所购买的商品的价格等。当控制装置向票据打印机2输出控制信号并对基于该印刷数据的印刷进行指示时,票据打印机2能够对表示购买记录的印刷物进行印刷。

[0027] 此外,控制装置能够对每个任意的期间内的购买记录进行汇总而取得销售额汇总,并且能够生成表示该销售额汇总的印刷数据。当控制装置向票据打印机2输出控制信号并对基于该印刷数据的印刷进行指示时,票据打印机2能够对表示店铺的销售额汇总的印刷物进行印刷。而且,控制装置能够对与在店铺工作的人员的出勤开始以及结束有关的信息进行记录并管理,且对每个任意的期间内的信息进行汇总并取得考勤信息。控制装置能够生成表示该考勤信息的印刷数据,且当向票据打印机2输出控制信号并对基于该印刷数据的印刷进行指示时,票据打印机2能够对表示在店铺工作的人员的考勤信息的印刷物进行印刷。

[0028] 如上所述,然票据打印机2能够对表示购买记录的印刷物进行印刷,也能够对作为其他的印刷物的表示销售额汇总的印刷物、表示考勤信息的印刷物进行印刷。另外,在本实施方式中,在被传送给票据打印机2的印刷数据中,包含有表示应该被印刷在印刷物上的文本的文本数据。

[0029] 在本实施方式中,票据打印机2具备能够向信息处理装置1发送印刷数据的发送

部。当实施印刷物的印刷指示时,票据打印机2基于印刷数据而实施由票据打印机2的印刷头(印刷部)进行的印刷,并且通过发送部的功能而向信息处理装置1发送印刷数据。另外,虽然在本实施方式中,票据打印机2与信息处理装置1进行连接,并从票据打印机2向信息处理装置1发送印刷数据,但也可以从对票据打印机2进行控制的控制装置向信息处理装置1发送印刷数据。在这种情况下,能够视为对票据打印机2进行控制的控制装置也构成票据打印机2。另外,发送部由通信接口、通信电路、通信端口等构成。此外,由发送部进行的印刷数据的发送既可以是由USB或以太网(イーサネット)等有线通信进行的发送,也可以是由Wi-Fi或Bluetooth等无线通信进行的发送(イーサネット、Wi-Fi、Bluetooth为注册商标)。

[0030] 信息处理装置1具备:处理器10(控制部)、HDD(硬盘驱动、存储部)20、通信部30。通信部30为,用于经由互联网等而与外部的设备进行通信的装置,且能够与票据打印机2实施通信并授受信息。另外,通信部30可以具有与票据打印机2的发送部同样的硬件结构,与发送部同样地,既可以为有线通信,也可以为无线通信。

[0031] 处理器10具备CPU、RAM等,且能够执行被记录在HDD20中的各种各样的程序。在本实施方式中,在这些程序中包含有信息处理程序11。信息处理程序11为,使处理器10执行对印刷数据中所包含的文本数据应用脚本的处理的程序。

[0032] 在本实施方式中,脚本为,对用于使在文本数据中所包含的特定的项目与表示项目的内容的参数建立对应的处理进行记述的程序。例如,在图5的左侧示出了表示购买记录的印刷物Pr。在该印刷物上,记载有印刷物被发行的日期时间(02/17/2017)、收款机的操作员姓名(Name A)、各个商品名的价格(\$4.99)等。该日期时间、操作员姓名、价格为可变的的信息,且在本实施方式中将这些称为参数。

[0033] 由于这些参数仅仅通过其信息本身(价格的值等)是毫无意义的,因此表示这些参数为与什么对应的参数的信息为项目。例如,由于4.99美元等这样的值本身而言其是什么的价格是不明确的,因此通过使其与商品名建立对应,从而具有特定的商品的价格为4.99美元这样的含义。此外,例如,由于2017年2月17日等这样的值本身而言其是什么日期时间是不明确的,因此通过使其与印刷物被发行的日期时间建立对应,从而具有发行日期时间为2017年2月17日这样的含义。

[0034] 如上文所述,在本实施方式中,在印刷物上印刷的文本通过参数(数值或者人名等的字符串)、和表示参数是什么参数的项目而被构成。因此,脚本被定义为,实施如下处理,即,对文本数据进行解析,且提取项目和参数并使两者建立对应。在本实施方式中,当实施该处理时生成使项目名与参数建立对应的文本数据。在此,将被生成的该文本数据称为结构化数据。另外,在结构化数据中,存在项目具有分层结构的情况。即,存在某个项目被进一步细化为下位的子项目的情况。当然,分层的数量也可以为三个以上。在图5所示的结构化数据中,通过对括号的种类进行区分,从而示出了这样的分层结构。

[0035] 在图5的右侧示出了从印刷物Pr被生成的结构化数据St的示例。在图5的右侧的结构化数据St中,项目名与参数通过“:”而被建立对应。例如,作为表示印刷物被发行的日期时间的参数的“2017-02-17T09:30:00.000Z”与“printed_at”这样的项目名建立对应。此外,作为表示对收款机进行了操作的操作员姓名的参数的“Name A”与“staff”这样的项目的子项目的项目名即“id”建立对应。另外,也可以在印刷物中不记载项目名。例如,虽然在图5的左侧所示的印刷物Pr中,印刷物被发行的日期时间用02/17/201709:30这样的数值被

记载,但是并未记载有表示其为发行日期时间的语句。

[0036] 在本实施方式中,如上所述,能够印刷三个种类的印刷物(表示购买记录的印刷物、表示销售额汇总的印刷物、表示考勤信息的印刷物),并且各种类中的项目和参数可以不同。因此,在本实施方式中,预先准备与印刷物的种类对应的脚本,并作为脚本数据20d而记录在HDD20中。

[0037] 在从票据打印机2发送印刷数据时,信息处理装置1对于在该印刷数据中被包含的文本数据应用脚本,而当根据印刷数据而被印刷的印刷物的种类不明确时,无法决定要应用的脚本。当在印刷物的种类不明确的状态下想要应用脚本时,则必须采用如下流程,即将所有的种类的脚本应用于文本数据,并采用适当的结果等。但是,当将所有的种类的脚本应用于文本数据时,处理将会花费时间。

[0038] 因此,在本实施方式中,采用了如下结构,即,在对印刷物的种类进行了判断的基础上,再决定应该应用的脚本。为了执行这样的处理,信息处理程序11具备取得部11a、判断部11b、应用部11c。即,当执行信息处理程序11时,处理器10作为取得部11a、判断部11b、应用部11c而发挥功能。

[0039] 通过取得部11a,而处理器10执行如下功能,即,从基于印刷数据而对印刷物进行印刷的票据打印机取得印刷数据的功能。即,当从票据打印机2发送印刷数据时,处理器10通过取得部11a的功能,而经由通信部30来取得印刷数据。当取得印刷数据时,作为印刷数据20a而被记录在HDD20中。

[0040] 通过判断部11b,而处理器10执行如下功能,即,基于被包含在印刷数据中的文本数据,对印刷物的种类进行判断的功能。在本实施方式中,处理器10通过基于使文本数据与印刷物的种类建立对应的示教数据而被实施了机器学习的学习模型,而对印刷物的种类进行判断。

[0041] 具体而言,本实施方式中的机器学习为,基于神经网络进行的学习。图2是示意性地表示在本实施方式中被利用的神经网络的结构图。在图2中,用圆圈表示模拟了神经元的节点,并用实线的直线表示节点间的结合(但是,为了将图简化,仅示出了节点以及结合的一部分)。此外,在图2中,将属于同一层的节点以在纵向上进行排列的方式来表示,左端的层为输入层Li、右端的层为输出层Lo。

[0042] 此处,在两层的关系中将靠近输入层Li的层称为上位层,将靠近输出层Lo的层称为下位层。即,成为如下的结构,即,在某个层中将上位1层的层的输出设为输入,并实施该输入和权重的相乘与偏差(bias)的加算,并经由激活函数而实施输出的结构。例如,在图2所示的层L_P的节点的数量为P个,对于层L_P的各个节点的输入数据为x₁...x_M,偏差为b₁...b_P,对于层L_P的第k个节点的权重为w_{k1}...w_{kM},激活函数为h(u)的情况下,该层L_P的各个节点处的中间输出u₁...u_P用以下的式1而被表现。

[0043] 数学式1

$$[0044] \quad (u_1 \cdots u_k \cdots u_P) = (x_1 \cdots x_k \cdots x_M) \begin{pmatrix} w_{11} & \cdots & w_{k1} & \cdots & w_{P1} \\ \vdots & & \vdots & & \vdots \\ w_{1k} & \cdots & w_{kk} & \cdots & w_{Pk} \\ \vdots & & \vdots & & \vdots \\ w_{1M} & \cdots & w_{kM} & \cdots & w_{PM} \end{pmatrix} + (b_1 \cdots b_k \cdots b_P) \quad \cdots \cdots (1)$$

[0045] 因此,若将式1代入激活函数 $h(u_k)$ 中,则可以获得层 L_P 的各个节点处的输出 $y_1 \cdots y_P$ 。

[0046] 模型通过以此方式被通用化的节点被设置多层从而被构成。因此,在本实施方式中,模型至少通过各层中的表示权重、偏差和激活函数的数据而被构成。另外,激活函数能够使用各种各样的函数,例如,S形(Sigmoid)函数、双曲函数(tanh)、ReLU函数(Rectified Linear Unit:线性整流函数)、其他的函数,优选为非线性函数。当然,模型中也可以包含有实施机器学习时需要的其他条件,例如,最优化算法的种类、学习率等参数。此外,层的数量、节点的数量、相关的结合关系能够设为各种各样的方式。

[0047] 在本实施方式中,构筑如下的模型,所述模型为,将被包含在印刷数据20a中的文本数据转换为预定的输入值并输入至输入层 L_i ,并且从输出层 L_o 输出表示印刷物的种类的输出值。在本实施方式中,构筑如下的模型,所述模型为,将表示在文本数据中是否包含有关键词的标志设为向输入层 L_i 输入的输入值、并将表示印刷物为特定的种类的可能性的值设为输出层 L_o 的输出值。例如,在图2中,在文本数据中存在“cash”(现金)这样的关键词的情况下,向节点 I_{i2} 输入数值1,在不存在“cash”这样的关键词的情况下,向节点 I_{i2} 输入数值0。此外,在文本数据中存在“grand total”(总计)这样的关键词的情况下,向节点 I_{iM} 输入数值1,在不存在“grand total”这样的关键词的情况下,向节点 I_{iM} 输入数值0。即,在图2所示的输入层 L_i 的各个节点中,一并记载有与各个节点对应的关键词。

[0048] 用于对关键词的有无进行判断的方法也可以为各种各样的方法,在本实施方式中,以特定的格式被记载的数值也能够成为关键词。例如,图2所示的“\$0.0”表示为,在表示美元货币的符号的右侧,排列了任意位数的任意的数值、小数点、任意位数的任意的数值的字符串,并且被设为在存在有该格式下的数值的情况下存在关键词。

[0049] 在本实施方式中,输出层 L_o 的节点 $T_{i1} \sim T_{i3}$ 为,根据印刷物的种类为三个种类的情况而被设置的三个节点,且节点 T_{i1} 、 T_{i2} 、 T_{i3} 分别对应于表示购买记录的印刷物、表示销售额汇总的印刷物、表示考勤信息的印刷物。即,在学习模型的学习完成之后,在文本数据为表示购买记录的印刷物的数据情况下,节点 T_{i1} 的输出值与其他的节点的输出值相比较为较大。在文本数据分别为表示销售额汇总的印刷物、表示考勤信息的印刷物的数据的情况下,节点 T_{i2} 、 T_{i3} 的输出值分别与其他节点的输出值相比较为较大。另外,在本实施方式中,以节点 T_{i1} 、 T_{i2} 、 T_{i3} 的输出值之和成为1的方式被标准化。因此,各个节点的输出的最大值为1。

[0050] 根据以上的模型,通过将表示被包含在文本数据中的关键词的有无的输入值输入至模型,且对各个节点的输出值进行计算,从而能够推断为,按照输出值从大到小的顺序,而作为与节点对应的种类的印刷物的可能性从高到低。

[0051] 虽然学习完成的模型作为学习模型20b而被记录在HDD20中,但是在基于被包含在印刷数据中的文本数据而对印刷物的种类进行推断之前,需要该学习模型20b被实施学习。该学习基于示教数据20c而被实施。即,示教数据20c为,使文本数据与根据该文本数据而被印刷的印刷物的所述种类建立对应的数据,并且为了进行学习而预先收集足够的量的数据。

[0052] 处理器10通过判断部11b的功能,从而在任意的时机基于示教数据20c而执行学习。此时,处理器10根据图2所示的模型,对模型进行调节以使根据示教数据20c而被建立了对应的文本数据与印刷物的种类之间的关系得以再现。即,处理器10分别基于被包含在示

教数据20c中的至少一部分的文本数据,而对关键词的有无进行确定,且将各个文本数据转换为输入值。

[0053] 而且,处理器10将该输入值输入至模型,并对模型进行调节以使所获得的输出值表示的印刷物的种类与示教数据20c表示的印刷物的种类一致(以使差分最小化)。通过该调节,在满足了被预先确定的学习完成判断基准的情况下,所获得的模型作为学习模型20b而被记录在HDD20中。

[0054] 处理器10在学习模型20b被记录在HDD20中的状态下实施印刷物的种类的判断。即,当取得印刷数据20a时,处理器10通过判断部11b的功能而从印刷数据20a中提取文本数据。此外,处理器10从该文本数据中对关键词进行检索,且对各个关键词的有无进行确定。另外,处理器10生成用标志(1或0)来表示关键词的有无的输入值,且基于学习模型20b而取得与输入值对应的输出值。而且,处理器10对各个节点的输出值进行比较,并按照输出层的值从大到小的顺序,视为作为与节点对应的印刷物的种类的可能性从高到低。

[0055] 通过以此方式对被印刷数据印刷的印刷物的种类为上述的三个种类中的各自的可能性被附加顺序,从而判断出印刷物的种类。当判断出印刷物的种类时,处理器10通过应用部11c的功能,从多个脚本中选择与印刷物的种类对应的脚本,且将选择的脚本应用于文本数据。

[0056] 在本实施方式中,由于确定了作为被印刷数据印刷的印刷物的种类的可能性从高到低的顺序,因此处理器10以越为可能性高的种类则使等级越高的方式,对选择与各个印刷物的种类相对应的脚本的优先等级进行设定。而且,处理器10根据所设定的优先等级,来选择并应用脚本,直到将所选择的脚本应用于文本数据并获得结果为止。

[0057] 即,本实施方式中的脚本为,对特定的格式的文本进行检测、或对特定的关键词与项目名的组合进行检测,并且针对于特定的参数而使与参数相对应的项目建立对应的处理的次序。在应用了脚本的情况下所获得的结果满足了既定的判断基准的情况下,处理器10判断为获得了脚本的应用结果。虽然既定的判断基准可以通过各种各样的方法来确定,但是在本实施方式中,在通过脚本的应用而获得的结构化数据中是否获得了被预先设为必须的项目的参数,成为判断基准。

[0058] 即,在被设为必须的项目的参数作为结构化数据而被取得了的情况下,处理器10判断为获得了脚本的结果。在被设为必须的项目的参数作为结构化数据而未被取得的情况下,处理器10判断为未获得脚本的结果。当然,被设为必须的项目也可以为多个。

[0059] 处理器10根据优先等级而实施这样的脚本的应用和结果是否被获得的判断。即,处理器10参照脚本数据20d,并将与优先等级被设为最高的印刷物的种类对应的脚本应用于印刷数据20a的文本数据,并取得结构化数据。处理器10基于该结构化数据而对是否获得了脚本的结果进行判断,且在获得了结果的情况下,处理器10结束脚本的应用。

[0060] 另一方面,在未获得结果的情况下,处理器10将下一个优先等级的脚本应用于印刷数据20a的文本数据,且基于结构化数据而对是否获得了结果进行判断。处理器10在之后直至判断为获得了结果为止,反复进行依次将优先等级下降1个等级而应用脚本并对是否获得了结果进行判断的处理。

[0061] 根据以上这样的本实施方式,处理器10基于被包含在印刷数据20a中的文本数据,而对与印刷物的种类对应的脚本进行选择并应用。此外,在本实施方式中,按照作为通过文

本数据而被印刷的印刷物的种类的可能性从高到低的顺序来应用脚本。其结果为,能够提高即使不应用所有的种类的脚本也能够将恰当脚应用于文本数据的可能性。

[0062] 另外,当通过以上的处理而根据印刷数据20a获得结构化数据时,该数据作为结构化数据20e而被记录在HDD20中,结构化数据20e能够利用于各种的解析。例如,多个印刷数据20a的结构化数据20e能够被提供给ISV(independent software vendor:独立软件供应商)等,从而被利用于各种解析和软件的开发等。

[0063] (2) 机器学习处理:

[0064] 接下来,对处理器10所执行的机器学习处理进行说明。图3为表示机器学习处理的流程图。机器学习处理在对印刷物的种类进行判断之前被预先执行。当开始机器学习处理时,处理器10通过判断部11b的功能,而取得训练模型(步骤S100)。即,取得被预先确定的神经网络的参数(权重、偏差等)作为训练模型。

[0065] 虽然在训练模型中,将文本数据中的关键词的有无设为输入值,且输出表示印刷物的种类的输出值,但是输出值所表示的印刷物的种类在初始并不准确。即,在训练模型中,虽然节点所构成的层的数量、节点的数量被确定,但是对输入输出的关系进行规定的参数(上述的权重、偏差等)并未被最优化。这些参数将在机器学习的过程中被最优化(即,被训练)。虽然在本实施方式中,训练模型被预先决定,但是利用者也可以对鼠标3a、键盘3b进行操作而决定。

[0066] 接下来,处理器10通过判断部11b的功能,而取得示教数据20c(步骤S105)。即,处理器10取得被预先记录在HDD20中的示教数据20c。另外,示教数据20c也可以在票据打印机2的运用过程、或开始新的格式的印刷物的使用之前等适当地被追加。当然,也可以删除根据已有的格式的印刷物的作废等而被作废的印刷物的示教数据20c。

[0067] 接下来,处理器10通过判断部11b的功能,而取得测试数据(步骤S110)。在本实施方式中,示教数据20c的一部分被设为测试数据。测试数据与示教数据20c被区分开。另外,虽然测试数据的信息量和示教数据20c的信息量可以为各种各样的量,但是在本实施方式中,以使示教数据20c与测试数据相比而更多的方式进行设定。

[0068] 接下来,处理器10通过判断部11b的功能,来决定初始值(步骤S115)。即,处理器10对在步骤S100中所取得的训练模型中的、可变的参数赋予初始值。初始值可以通过各种各样的方法而被决定。例如,能够将随机值或0等设为初始值,也可以出于根据权重和偏差而有所不同的思想来决定初始值。当然,也可以在学习的过程中以使参数被最优化的方式对初始值进行调节。

[0069] 接下来,处理器10通过判断部11b的功能,而实施学习(步骤S120)。即,处理器10提取示教数据20c的各个文本数据,且基于各个文本数据中的关键词的有无而将各个文本数据转换为输入值。而且,处理器10向在步骤S100中取得的训练模型输入该输入值,并且按照每个印刷物的种类,而对表示作为各个种类的可能性的输出值进行计算。此外,处理器10通过损失函数而对误差进行确定,所述损失函数表示输出值与表示和各个文本数据建立了对应的印刷物的种类的标志之差。而且,处理器10以既定次数反复进行基于由损失函数的参数进行的微分而对参数进行更新的处理。

[0070] 在将被包含在示教数据20c中的采样的数用*i*来表现的情况下,处理器10通过将根据第*i*个采样的文本数据而被确定的输入值输入至训练模型,从而取得第*i*个采样的输出值

($T_{i1} \sim T_{i3}$)。另一方面,表示与示教数据20c建立了对应的印刷物的种类的信息为,与印刷物的种类对应的值($t_{i1} \sim t_{i3}$),与特定的印刷物的种类对应的值为1,与其他的种类对应的值为0。

[0071] 因此,当将训练模型的输出值记载为 T_i ,且将示教数据20c所表示的印刷物的种类记载为 t_i 时,关于第 i 个采样的损失函数可以记载为 $L(T_i, t_i)$ 。当然,损失函数能够采用各种各样的函数,例如,能够采用交叉熵误差等。对以上那样的损失函数 L 进行计算的处理,针对于示教数据20c所表示的采样的全部或者一部分而被实施,且通过其平均或总和来表现1次的学习中的损失函数。例如,在通过总和来表现损失函数的情况下,全体的损失函数 E 通过接下来的式2而被表现。

[0072] 数学式2

$$[0073] \quad E = \sum_i L_i(T_i, t_i) \quad \cdots \cdots (2)$$

[0074] 获得了损失函数 E 之后,处理器10通过既定的最优化算法,例如,随机梯度下降法等而对参数进行更新。

[0075] 通过采用如上所述的方式,当实施了既定次数的参数的更新时,处理器10对训练模型的泛化是否完成进行判断(步骤S125)。即,处理器10将根据在步骤S110中所取得的测试数据而被计算出的输入值输入至训练模型并取得输出值。而且,处理器10通过取得与最大的输出值对应的印刷物的种类和与测试数据建立了对应的印刷物的种类相一致的数,且将其除以测试数据的采样数,从而取得推断精度。在本实施方式中,处理器10在推断精度为阈值以上的情况下判断为泛化完成。

[0076] 另外,除了泛化性能的评估以外,也可以实施超参数的合理性的验证。即,在作为权重和偏差以外的可变量的超参数、例如节点的数等被调谐的结构中,处理器10也可以基于验证数据而对超参数的合理性进行验证。验证数据通过例如从示教数据20c中被预先提取,并与示教数据20c和测试数据区分而被准备。

[0077] 在步骤S125中,在未判断为训练模型的泛化完成的情况下,处理器10反复进行步骤S120。即,进一步对权重以及偏差进行更新。另一方面,在步骤S125中,在判断为训练模型的泛化完成的情况下,处理器10对学习模型进行记录(步骤S130)。即,处理器10将训练模型作为学习模型20b而记录在HDD20中。根据通过以此方式获得的学习模型20b而对印刷物的种类进行推断的结构,能够基于被包含在任意的印刷数据中的文本数据,而对通过印刷数据而实施印刷的印刷物的种类进行推断。

[0078] (3) 信息处理:

[0079] 接下来,对处理器10所执行的信息处理进行说明。图4是表示信息处理的流程图。在本实施方式中,信息处理在从任意的票据打印机2发送了印刷数据的情况下被执行。

[0080] 当开始信息处理时,处理器10通过取得部11a的功能,而取得印刷数据(步骤S200)。即,处理器10通过取得部11a的功能,经由通信部30而与票据打印机2进行通信,并接收票据打印机2发送的印刷数据。处理器10将接收到的印刷数据作为印刷数据20a而记录在HDD20中。

[0081] 接下来,处理器10通过判断部11b的功能而实施关键词检索(步骤S205)。即,处理器10从在步骤S200中所接收到的印刷数据20a中提取文本数据,并且对是否存在被预先确

定的关键词进行判断。处理器10将1与存在关键词建立对应,并将0与不存在关键词建立对应。例如,若为图2所示的示例,则实施如下处理,即,在存在由“\$0.0”表示的格式的关键词的情况下将标志设定为1,并在存在“cash”这样的关键词的情况下将标志设定为1,并在不存在“hours”(小时)这样的关键词的情况下将标志设定为0等的处理。

[0082] 接下来,处理器10实施由学习模型进行的推断(步骤S210)。即,处理器10基于学习模型20b,而实施将在步骤S205中被设定的标志作为输入值的神经网络的运算。其结果为,能够获得关于与印刷物的种类对应的各个节点的输出值。例如,在图2所示的示例中,当将在步骤S205中被设定的标志设为输入值,且实施由学习模型20b进行的运算时,关于各个节点 T_{i1} 、 T_{i2} 、 T_{i3} 能够分别获得0以上且1以下的输出值。

[0083] 接下来,处理器10取得优先等级(步骤S215)。即,处理器10视为各个节点的输出值的大小表示作为与各个节点对应的种类的印刷物的可能性。而且,处理器10按照输出值的从大到小的顺序来排列与各个节点对应的种类,并设为在应用与该种类对应的脚本时的优先等级。

[0084] 例如,在图2所示的示例中,节点 T_{i1} 、 T_{i2} 、 T_{i3} 分别对应于表示购买记录的印刷物、表示销售额汇总的印刷物、表示考勤信息的印刷物。在该示例中,在节点 T_{i1} 、 T_{i2} 、 T_{i3} 的输出值为0.1、0.85、0.05的情况下,成为如下的优先等级,即,与表示销售额汇总的印刷物对应的脚本的优先等级为最高,接下来是与表示购买记录的印刷物对应的脚本,再下来是与表示考勤信息的印刷物对应的脚本。

[0085] 接下来,处理器10对未应用且优先等级最高的脚本进行应用(步骤S220),且直至被判断为获得了结果为止反复进行步骤S220(步骤S225)。即,处理器10从在步骤S200中所取得的印刷数据中提取文本数据,且按照在步骤S215中取得的优先等级从高到低的顺序将脚本应用于文本数据。当脚本被应用于文本数据时,由于获得结构化数据,因此处理器10基于该结构化数据,而对是否满足既定的判断基准进行判断。而且,在满足既定的判断基准的情况下,处理器10判断为获得了结果。在步骤S225中,在判断为获得了结果的情况下,处理器10将所获得的结构化数据20e保存在HDD20中(步骤S230)。根据以上的结构,按照作为通过文本数据而被印刷的印刷物的种类的可能性从高到低的顺序来应用脚本。其结果为,能够提高即使不应用所有种类的脚本也能够将恰当脚本应用于文本数据的可能性。

[0086] (4) 其他的实施方式:

[0087] 以上的实施方式为用于实施本发明的一个示例,另外也可以采用各种各样的实施方式。例如,本发明的一个实施方式所涉及的信息处理装置既可以通过各种各样的方式来提供,也可以通过云服务器等多个装置来提供。而且,通过信息处理装置1而被执行的功能的至少一部分也可以通过票据打印机2来执行。例如也可以设为,学习模型20b被保存在票据打印机2中,且在票据打印机2中对印刷物的种类进行判断。另外,如以上的实施方式那样,基于被包含在印刷数据中的文本数据,而对通过印刷数据而被印刷的印刷物的种类进行判断的方法,也能够作为程序的发明、方法的发明来实现。

[0088] 另外,机器学习和印刷物的种类的判断也可以通过其他的装置而被执行。另外,上述的实施方式为一个示例,能够采用省略一部分的结构、或追加其他的结构的实施方式。

[0089] 取得部只要能够从基于印刷数据而对印刷物进行印刷的票据打印机,取得印刷数据即可。即,取得部只要能够在任意的时机取得用于使印刷任意的印刷物的印刷数据即可。

票据打印机与信息处理装置的连接方式既可以是各种各样的方式,也可以通过电缆而直接地进行连接,还可以经由互联网等而进行连接。

[0090] 票据打印机只要能够对票据进行印刷即可,当然,也可以对印刷票据以外的印刷物实施印刷。票据只要为表示在店铺的购买记录的印刷物即可,且只要对购买记录进行印刷,则印刷内容和布局不受限制。印刷数据只要为用于使票据打印机印刷任意的印刷物的数据即可,可以为任意的格式。但是,在印刷数据中包含有文本数据,并且该文本数据所表示的文本被印刷在印刷纸张上从而成为印刷物。当然,也可以在印刷物上印刷文本以外的对象物,在印刷数据中也可以包含有用于印刷各种图像、格线等的的数据。

[0091] 另外,文本数据只要表示被印刷在印刷物上的字符串即可,当然也可以包含文本的布局、字体的类型、大小等信息。这些信息也可以被利用于印刷物的种类的判断。

[0092] 判断部只要能够基于被包含在印刷数据中的文本数据,而对印刷物的种类进行判断即可。即,文本数据表示在印刷物上要被实施印刷的文本,且通过文本而对印刷物的内容进行记述。因此,能够将文本作为用于对印刷物的种类进行判断的要素而利用。用于判断印刷物的种类的方法并不限定于如上述的实施方式那样利用由机器学习进行的学习完成的模型的结构。例如,也可以采用如下结构等,即,根据在文本数据所表示的文本内包含有与印刷物的种类对应的的特定的关键词的比率等,来判断出印刷物的种类。

[0093] 印刷物的种类只要根据印刷物的用途等而被预先分类即可,除了如上述的实施方式那样,表示在店铺的购买记录的印刷物、表示在店铺的销售额汇总的印刷物、表示考勤信息的印刷物以外,还可以列举出各种各样的印刷物。例如,既可以为表示所购买的彩票、票券等的印刷物,也可以为表示公共设施的使用费的支付证明的印刷物,另外也可以为各种各样的印刷物。

[0094] 应用部只要能够从多个脚本中选择与印刷物的种类对应的脚本,并且将所选择的脚本应用于文本数据即可。即,印刷物的种类存在多个,且预先准备有与其分别相对应的脚本。虽然由于脚本根据印刷物的种类而被准备,因此至少存在有与印刷物的种类的数量相同的数量,但是也可以针对于一个种类的印刷物而预先准备多个脚本。

[0095] 脚本只要为用于基于文本数据并按照既定的顺序而实施处理的程序即可,语言不受限制。脚本为简单的处理的顺序的情况较多,当然,处理的内容也可以为复杂的。应用脚本的目的既可以为各种各样的目的,也可以针对于每个脚本而存在目的。作为目的,除了如上述的实施方式那样,将项目与参数建立对应并提取,且作为各种解析的原数据而进行提供的目的以外,可以设想各种各样的目的。例如,设想如下目的等等的目的,即,基于在购买记录中所包含的每个商品的金额或合计金额等而对用于进行家庭收支管理的信息进行收集的目的、基于店铺的销售额汇总而对用于进行市场调查的信息进行收集的目的、基于考勤信息而对用于进行劳务管理的信息进行收集的目的等。

[0096] 脚本对于文本数据的应用,相当于执行根据脚本所表示的顺序而对文本数据进行解析的处理。在此,只要基于文本数据而执行既定的处理即可,脚本既可以是翻译器型也可以是编译器型。

[0097] 示教数据只要使被包含在印刷数据中的文本数据与通过该印刷数据而被印刷的印刷物的种类建立对应即可。示教数据只要以与学习模型的输入值以及输出值的格式相同的格式被定义即可,并不限定于如上述的实施方式那样的格式。既可以将文本数据所表示

的字符串设为输入值,也可以对表示在文本数据中能被包含的字符串的one-hot (独热) 向量进行定义并设为输入值,还可以将被包含在文本数据中的单词或单词的数量设为输入值,从而能够采用各种各样的结构。

[0098] 在从文本数据中被导出的值成为输入值的情况下,该值的导出也可以通过各种各样的方法而被实施。例如,也可以对字符串或单词进行检索,而取得检索结果的有无或被检索到的数量。另外,也可以对货币、时刻是否以特定的格式而被表现进行判断。例如,也可以对如在字符串的左侧存在表示货币的记号、在其右侧存在位数的数值、在其右侧存在小数点、在其右侧还存在数值那样的字符串进行检索来确定价格的有无。当然,在这样的检索时,也可以利用正则表达式。

[0099] 此外,输出值既可以针对于印刷物的多个种类,而表示作为每个种类的可能性,也可以表示并非为每个种类的可能性。此外,既可以针对于印刷物的每个种类而生成学习模型,并且针对印刷物的每个种类而表示作为该种类的可能性的值被设为输出值,另外可以设想各种各样的方式。当然,输出值也可以表示脚本。

[0100] 机器学习的方法也可以为各种各样的方法。即,只要以如下方式进行机器学习即可,即,构筑将文本数据本身或从文本数据中被导出的值(单词的数或文本的长度、空格数、人名的有无等)输入,并输出印刷物的种类的模型,并且通过该模型而获得恰当的输出。在机器学习中,只要能够进行使由该模型进行的输出与示教数据中的输出值的差分极小化的学习即可。

[0101] 因此,例如,在实施由神经网络进行的机器学习的情况下,只要适当地对如下要素等各种各样的要素进行选择并实施机器学习即可,所述要素为,构成模型的层的数量和节点的数量、激活函数的种类、损失函数的种类、梯度下降法的种类、梯度下降法的最优化算法的种类、小批量学习的有无和批量的数量、学习率、初始值、过度学习抑制方法的种类和有无、卷积层的有无、卷积运算中的过滤器的大小、过滤器的种类、填充(padding)和跨度(stride)的种类、池化(pooling)层的种类和有无、全耦合层的有无、递归结构的有无等要素。当然,也可以通过其他的机器学习、例如支持向量机(Support vector machine)和聚类(clustering)、强化学习等而被实施学习。

[0102] 另外,也可以实施模型的结构(例如,层的数量、每层的节点的数量等)自动被最优化的机器学习。另外,在信息处理装置中被实施机器学习的结构中,也可以为从多个票据打印机收集示教数据,并基于该示教数据而实施机器学习的结构。

[0103] 当然,文本数据也可以通过RNN(Recurrent Neural Network:循环神经网络)而实施处理。即,由于如果根据RNN,则能够考虑到文本数据的前后关系,因此也可以利用包含由RNN进行的处理的模型的机器学习。图6为包含RNN的模型的结构例。在图6中,成为输入值的文本数据被分解为单词,且被设为输入值。在图6中,示出了如下示例,即,作为构成购买记录的文本的一部分的“Cashier NameA Milk\$4.99、...、0.70”这样的连续的字符串被分解为单词,“Cashier”(收银员)被设为输入值 x_1 、“Name A”(姓名A)被设为输入值 x_2 、“Milk”(牛奶)被设为输入值 x_3 、“\$”被设为输入值 x_4 、“4.99”被设为输入值 x_5 ,之后,“0.79”被转换为输入值 x_n 的示例。

[0104] 另外,在图6中,设想了各输入值被输入至RNN的模块 $M_1 \sim M_5$ 中并输出输出值 $y_1 \sim y_5$ 、...、 y_n 的示例。另外,在图6所示的示例中,RNN的模块 $M_1 \sim M_5$ 、...、 M_{n-1} 的输出值 $y_1 \sim y_5$ 、...、 y_{n-1}

被输入至下一个模块 $M_2 \sim M_6$ 、 \dots 、 M_n 中。另外,与输入值 x_1 一起被输入至RNN的第一个模块 M_1 的值 y_0 ,可以通过各种各样的方法而被初始化(例如,0向量等)。在图6所示的示例中,这些输出值被输入至神经网络(NN)中,并最终获得与印刷物的种类对应的每个节点 T_{i1} 、 T_{i2} 、 T_{i3} 的输出值。

[0105] 在以上那样的模型中,输入值的定义法、权重、偏差等的RNN的模型或NN的模型,能够利用公知的各种各样的方法。如果在图1所示的示例中,设想以上那样的模型,且处理器10通过判断部11b的功能而实施机器学习,则能够对输入文本数据,并对通过印刷数据而被印刷的印刷物的种类进行推断的模型进行机器学习。而且,即使在这样的结构中,也能够通过对印刷物的种类进行判断,且应用与判断结果相应的脚本,从而能够提高即使不应用所有的种类的脚本也能够将恰当的脚本应用于文本数据的可能性。当然,图6所示的RNN的模型也为一个示例,也可以采用双方向RNN等各种各样的模型。

[0106] 印刷物作为多个种类中的每一种的可能性,也可以通过各种各样的方法而被定义。因此,除了印刷物作为特定的种类的可能性的最大值为1、最小值为0的上述的实施方式以外,也可以采用各种各样的结构。无论采用哪个结构,只要能够通过表示印刷物为特定的种类的数值的大小关系,而对通过印刷数据而被印刷的印刷物为多个印刷物中的每一个的可能性附加等级即可。只要优先等级通过该附加等级而被确定即可。

[0107] 将所选择的脚本应用于文本数据并且是否获得了结果,只要通过被预先确定的判断基准而被进行判断既可。因此,如上述的实施方式那样在脚本输出既定的结构化数据的情况下,除了在获得了在结构化数据之中被预先设为必要的数据的情况下视为获得了结果的结构以外,也可以采用各种各样的结构。例如可以列举出如下结构等,即,在应用了脚本的情况下,在错误的发生比率、或项目与参数之间的对应关系成为了不确定的比率等为阈值以下的情况等下,判断为获得了脚本的应用结果。

[0108] 另外,本发明也能够作为计算机所执行的程序、方法而应用。此外,如上所示的程序、方法既有作为单独的装置而被实现的情况,也有利用多个装置所具备的部件而被实现的情况,从而包含各种的方式。此外,能够设为一部分为软件且一部分为硬件等、适当地进行变更。另外,即使作为程序的记录介质,发明也成立。当然,该程序的记录介质既可以为磁记录介质也可以为半导体存储器等,且即使在今后开发的任何的记录介质中,也能够被认为是完全相同的。

[0109] 符号说明

[0110] 1...信息处理装置;2...票据打印机;3a...鼠标;3b...键盘;10...处理器;11...信息处理程序;11a...取得部;11b...判断部;11c...应用部;20...HDD;20a...印刷数据;20b...学习模型;20c...示教数据;20d...脚本数据;20e...结构化数据;30...通信部。

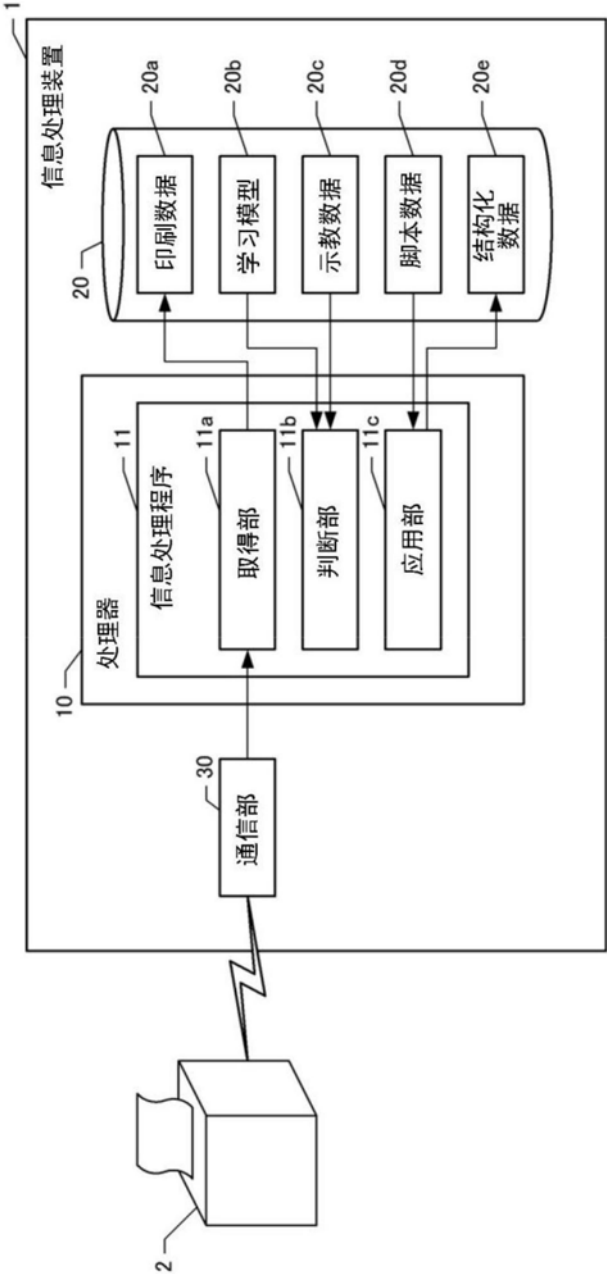


图1

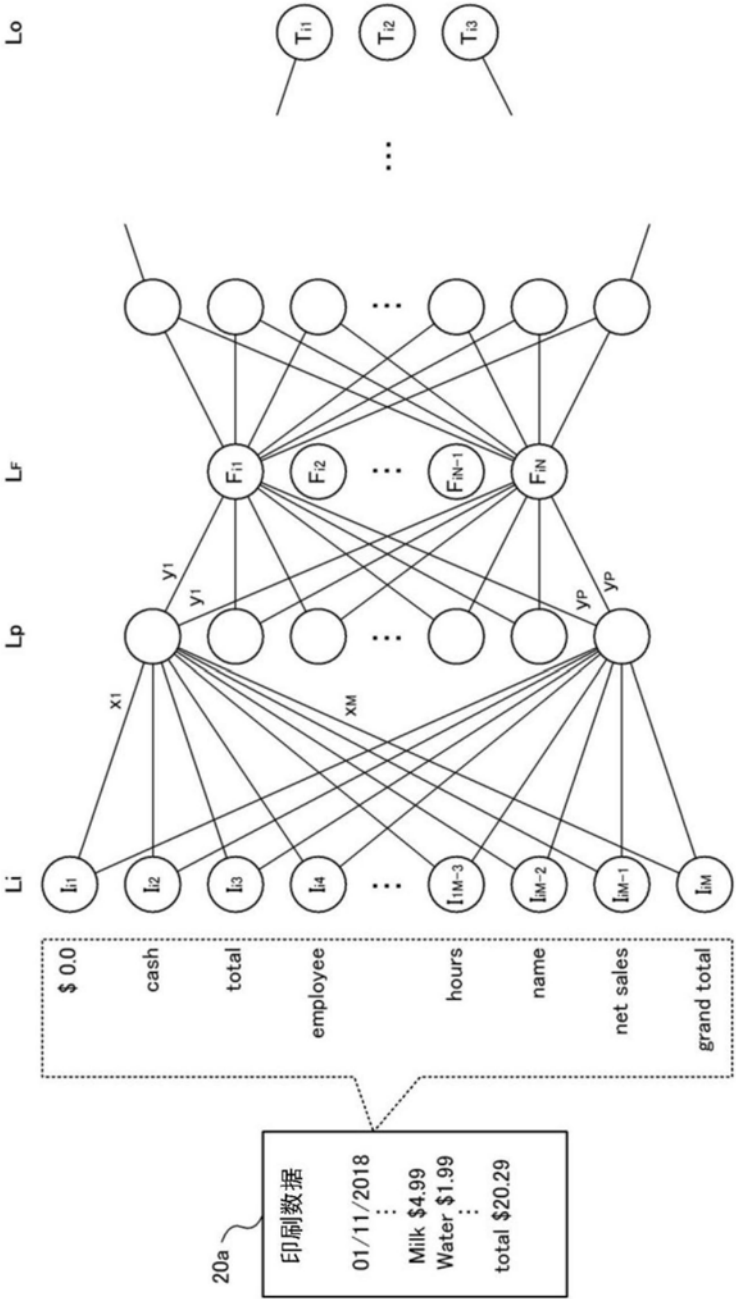


图2

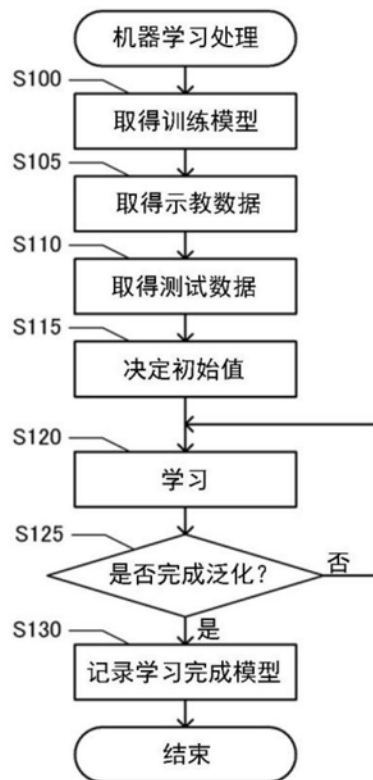


图3

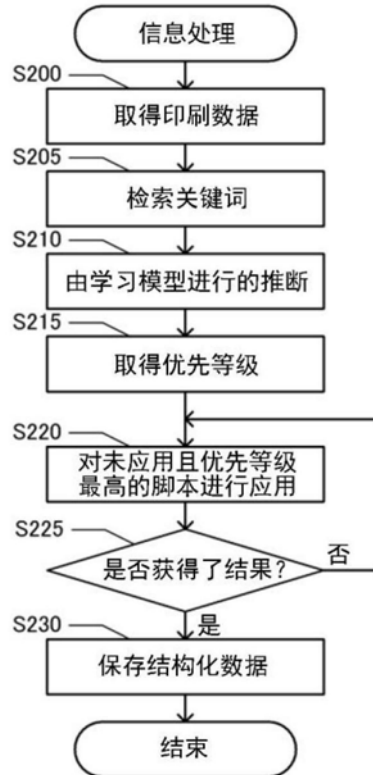


图4

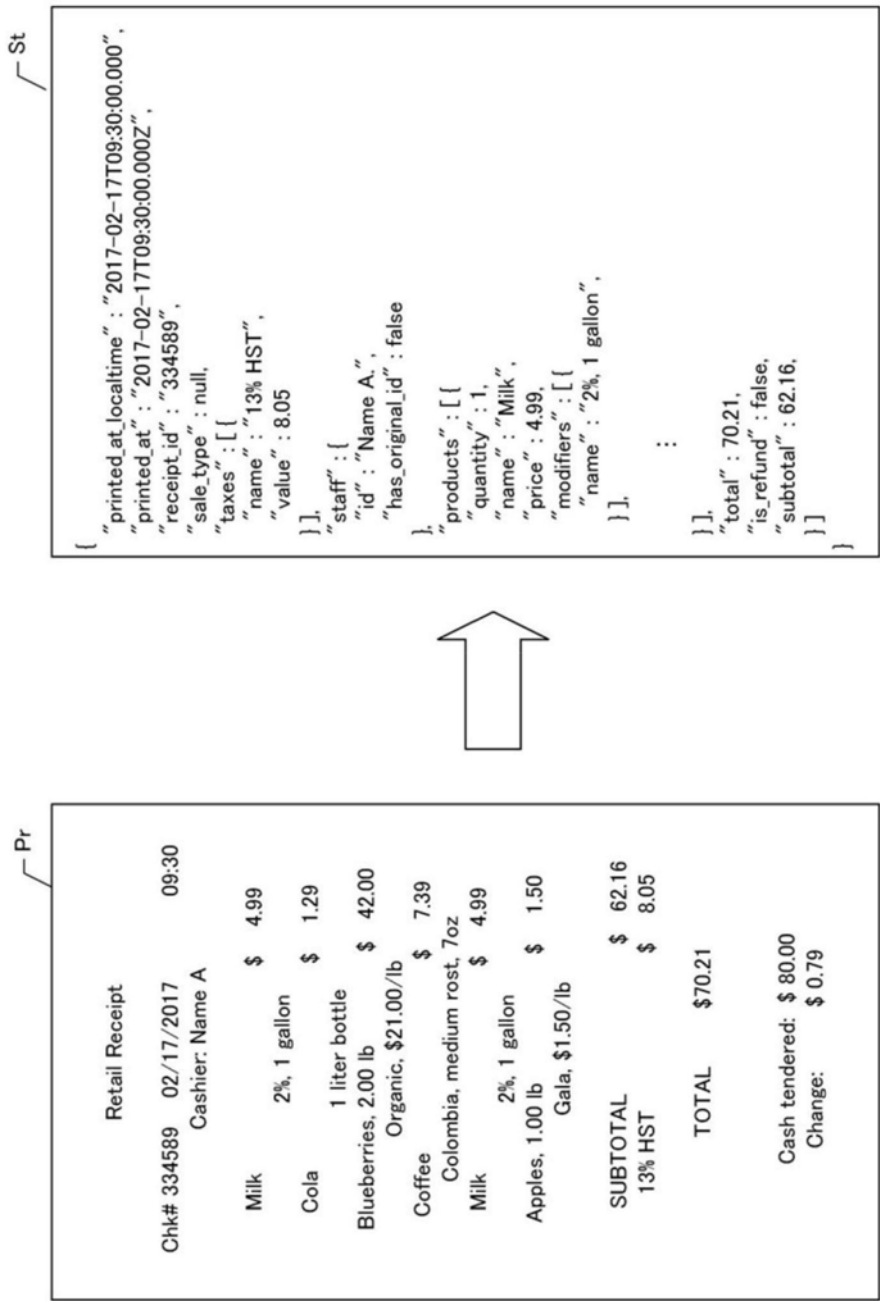


图5

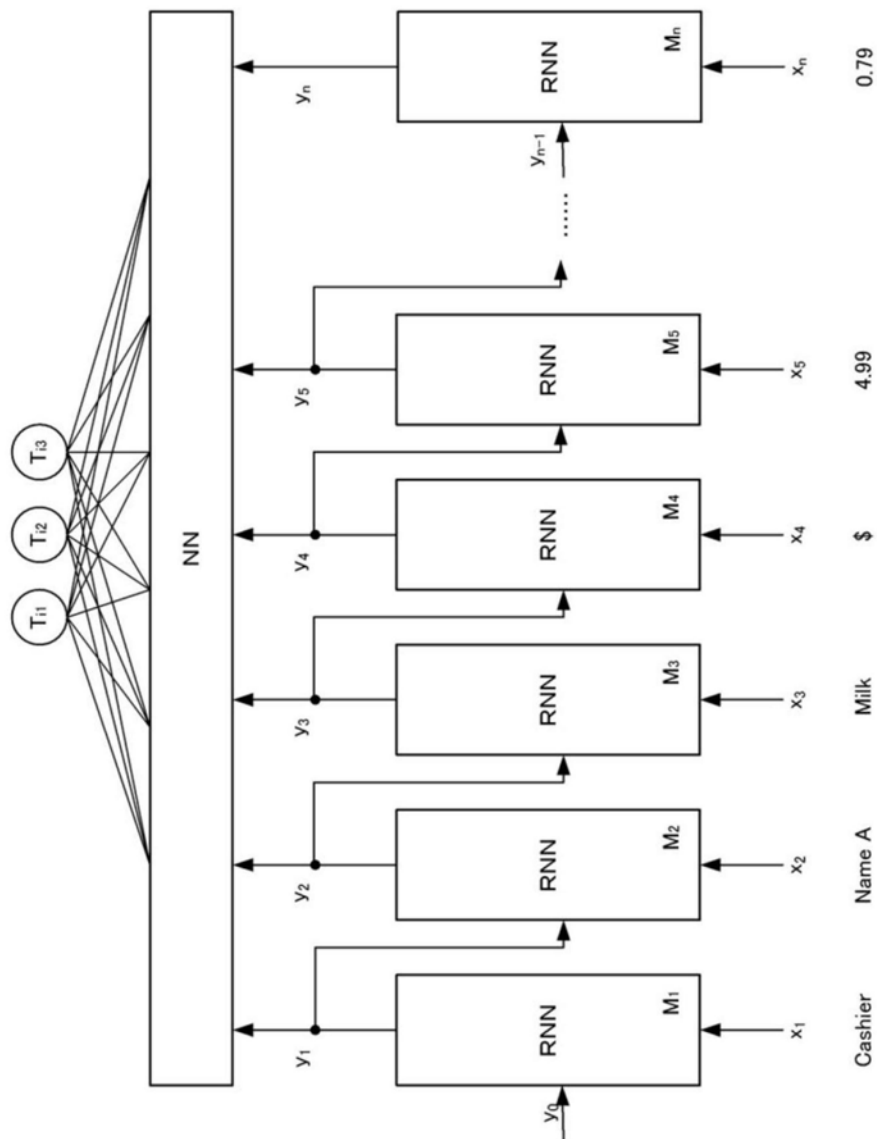


图6