



(19) 中華民國智慧財產局

(12) 發明說明書公告本

(11) 證書號數：TW I724515 B

(45) 公告日：中華民國 110 (2021) 年 04 月 11 日

(21) 申請案號：108130654

(22) 申請日：中華民國 108 (2019) 年 08 月 27 日

(51) Int. Cl. : G06N20/00 (2019.01)

G06N5/00 (2006.01)

(71) 申請人：聯智科創有限公司 (中華民國) UNITED SMART ELECTRONICS CORPORATION
(TW)

臺南市永康區永科三路 99 號 3 樓

(72) 發明人：許智威 HSU, CHIH-WEI (TW)；李佳穎 LI, JIA-YING (TW)；李奇 LI, CHI (TW)；
江啓睿 JIANG, CI-RUEI (TW)

(74) 代理人：高玉駿；楊祺雄

(56) 參考文獻：

TW 201829982A

CN 102326144A

CN 104598917A

CN 106228240A

US 2019/0149166A1

審查人員：陳守德

申請專利範圍項數：9 項 圖式數：4 共 19 頁

(54) 名稱

機器學習服務提供方法

(57) 摘要

一種機器學習服務提供方法，包含以下步驟：(A)藉由一處理裝置根據訓練資料集，自動判定出一適用該訓練資料集的機器學習演算法；(B)藉由該處理裝置根據該訓練資料集，利用該機器學習演算法，進行訓練以獲得一機器學習推論模型；(C)藉由該處理裝置將訓練完的該機器學習推論模型，利用一轉換套件，轉換成一至少在一可程式化電路裝置上被運行的執行檔案；及(D)藉由一電連接該可程式化電路裝置的處理單元，將一待分析資料傳送至運行有該執行檔案的該可程式化電路裝置，以使該可程式化電路裝置進行推論運算以獲得一推論結果。

指定代表圖：

符號簡單說明：

21~25 步驟

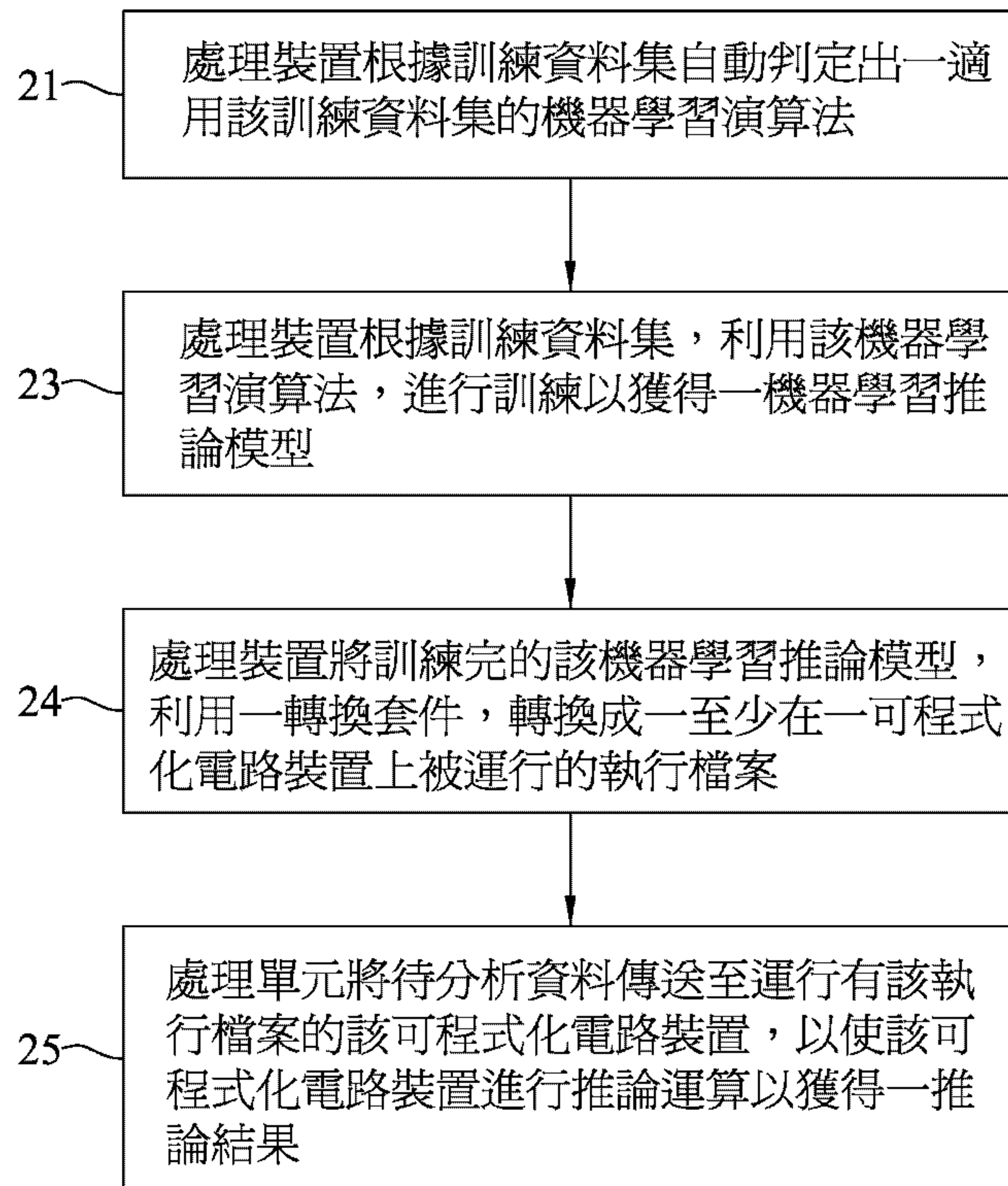


圖2



公告本

【發明摘要】

I724515

【中文發明名稱】 機器學習服務提供方法

【中文】

一種機器學習服務提供方法，包含以下步驟：(A)藉由一處理裝置根據訓練資料集，自動判定出一適用該訓練資料集的機器學習演算法；(B)藉由該處理裝置根據該訓練資料集，利用該機器學習演算法，進行訓練以獲得一機器學習推論模型；(C)藉由該處理裝置將訓練完的該機器學習推論模型，利用一轉換套件，轉換成一至少在一可程式化電路裝置上被運行的執行檔案；及(D)藉由一電連接該可程式化電路裝置的處理單元，將一待分析資料傳送至運行有該執行檔案的該可程式化電路裝置，以使該可程式化電路裝置進行推論運算以獲得一推論結果。

【指定代表圖】：圖(2)。

【代表圖之符號簡單說明】

21~25..步驟

【發明說明書】

【中文發明名稱】 機器學習服務提供方法

【技術領域】

【0001】本發明是有關於一種服務提供方法，特別是指一種提供機器學習服務的機器學習服務提供方法。

【先前技術】

【0002】近年來，機器學習的快速發展，應用層面遍及醫療、經濟及工業等各領域，許多廠商紛紛陸續將機器學習技術引入公司的產品或製程。然而，受限於硬體能力與機器學習演算之佈署研發的困難度，而無法有效地運用機器學習技術。

【0003】因此，有許多公司致力於發展人工智慧軟體平台(，亦即AI平台)，以提供廠商應用AI平台來快速建構自己的人工智慧產品，而在硬體方面，也有許多公司研發人工智慧加速晶片以及適用於人工智慧的嵌入式裝置，來加速機器學習的執行效率。因此，如何提供一套有效加速機器學習執行效率與降低使用複雜度的機器學習服務，以幫助使用者有效地運用機器學習技術，便成為一亟待解決的問題。

【發明內容】

第1頁，共10頁(發明說明書)

【0004】因此，本發明的目的，即在提供一種加速機器學習執行效率與降低使用複雜度的機器學習服務提供方法。

【0005】於是，本發明機器學習服務提供方法，包含以下步驟：

【0006】(A)藉由一處理裝置根據訓練資料集，自動判定出一適用該訓練資料集的機器學習演算法；

【0007】(B)藉由該處理裝置根據該訓練資料集，利用該機器學習演算法，進行訓練以獲得一機器學習推論模型；

【0008】(C)藉由該處理裝置將訓練完的該機器學習推論模型，利用一轉換套件，轉換成一至少在一可程式化電路裝置上被運行的執行檔案；及

【0009】(D)藉由一電連接該可程式化電路裝置的處理單元，將一待分析資料傳送至運行有該執行檔案的該可程式化電路裝置，以使該可程式化電路裝置進行推論運算以獲得一推論結果。

【0010】本發明的功效在於：藉由該處理裝置根據該訓練資料集，自動判定出適用該訓練資料集的該機器學習演算法，藉此使用者僅需提供該訓練資料集，該處理裝置即會自動判定出適用該訓練資料集的機器學習演算法，進而可大幅降低使用者判定使用何種機器學習方法的負擔，也無需自行設計用於該訓練資料集的機器學習演算法。接著，藉由該處理裝置利用該機器學習演算法訓練出該機器學習推論模型，繼而，將訓練完的該機器學習推論模型，轉換成

至少在該可程式化電路裝置上被運行的該執行檔案，藉由該可程式化電路裝置來進行推論運算，可加速機器學習執行效率。

【圖式簡單說明】

【0011】本發明的其他的特徵及功效，將於參照圖式的實施方式中清楚地呈現，其中：

圖 1 是一方塊圖，說明一用於執行本發明機器學習服務提供方法的實施例的機器學習服務提供系統；

圖 2 是一流程圖，說明本發明機器學習服務提供方法的實施例；

圖 3 是一流程圖，說明當所欲訓練出之該機器學習推論模型屬於一預測模型時，要採用何種演算法；及

圖 4 是一流程圖，說明當所欲訓練出之該機器學習推論模型屬於一分類模型時，要採用何種演算法。

【實施方式】

【0012】參閱圖 1 與圖 2，本發明機器學習服務提供方法的實施例，藉由一機器學習服務提供系統 1 來實施。該機器學習服務提供系統 1 包含一處理裝置 11、一可程式化電路裝置 12，及一電連接該可程式化電路裝置 12 的處理單元 13。

【0013】在本實施例中，該處理裝置11例如為一個人電腦、一伺服器、一筆記型電腦或一平板電腦，但不限於此。該可程式化電路裝置12例如為一可重複燒錄及快速編寫的現場可程式化邏輯閘陣列(Field Programmable Gate Array，簡稱FPGA)。該處理單元13例如為一中央處理單元13(Central Processing Unit，簡稱CPU)，該處理單元13與該處理裝置11的中央處理單元13可為同一處理單元13，也可以是獨立的兩個處理單元13。

【0014】參閱圖1、圖2與圖3，本發明機器學習服務提供方法的實施例包含以下步驟。

【0015】在步驟21中，該處理裝置11根據訓練資料集，自動判定出一適用該訓練資料集的機器學習演算法。其中，該訓練資料集包含建模資料與驗證資料，且該機器學習演算法包含一類神經演算法、一支持向量機演算法及一k-平均演算法等之其中一者。在本實施例中，該訓練資料集包含多筆訓練資料，每筆訓練資料皆包括輸入特徵及對應於該輸入特徵的正確結果(亦即，標籤)，然而，在其他實施方式中，每筆訓練資料亦可僅包括輸入特徵，但不包括對應於該輸入特徵的正確結果。值得一提的是，該處理裝置11預先實作了各種不同的機器學習演算法如，該類神經演算法、該支持向量機演算法及該k-平均演算法等，該處理裝置11可根據該訓練資料集判定出適用該訓練資料集的該機器學習演算法。

【0016】又值得一提的是，步驟21包含以下子步驟。

【0017】在子步驟211中，該處理裝置11根據該訓練資料集判定所欲訓練出之該機器學習推論模型是屬於一預測模型或一分類模型。當該處理裝置11判定出所欲訓練出之該機器學習推論模型屬於該預測模型時，進行子步驟212；當該處理裝置11判定出所欲訓練出之該機器學習推論模型屬於該分類模型時，進行子步驟217。在本實施例中，該處理裝置11係根據該訓練資料集之標籤來判定所欲訓練出之該機器學習推論模型是屬於該預測模型或該分類模型，當該訓練資料集具有標籤且其標籤之屬性不屬於類別而是連續的值時，該處理裝置11判定出所欲訓練出之該機器學習推論模型屬於該預測模型；而當該訓練資料集不具有標籤，或具有標籤且其標籤之屬性屬於類別時，該處理裝置11判定出所欲訓練出之該機器學習推論模型屬於該分類模型。

【0018】在子步驟212中，該處理裝置11根據該訓練資料集判定該訓練資料集之輸入特徵是否為文字型態。當該處理裝置11判定出該訓練資料集之輸入特徵為文字型態時，進行子步驟213；當該處理裝置11判定出該訓練資料集之輸入特徵非為文字型態時，進行子步驟214。

【0019】在子步驟213中，該處理裝置11將該支持向量機演算法作為該機器學習演算法。

【0020】在子步驟214中，該處理裝置11根據該訓練資料集中的建模資料，分別利用該支持向量機演算法與該類神經演算法進行訓練以分別獲得一第一推論模型與一第二推論模型。

【0021】在子步驟215中，該處理裝置11根據該訓練資料集中的驗證資料，分別利用該第一推論模型與該第二推論模型進行推論以分別獲得一第一推論準確度與一第二推論準確度。

【0022】在子步驟216中，該處理裝置11將具有較高推論準確度之推論模型所對應的演算法作為該機器學習演算法。

【0023】值得一提的是，藉由該處理單元13分別利用該支持向量機演算法與該類神經演算法訓練出該第一推論模型與該第二推論模型，並將具有較高推論準確度之推論模型所對應的演算法作為該機器學習演算法，可使得所獲得之機器學習演算法不僅適用該訓練資料集，利用該機器學習演算法所獲得的機器學習推論模型還是準確度較高的推論模型。

【0024】在子步驟217中，該處理裝置11根據該訓練資料集判定該訓練資料集是否包含標籤。當該處理裝置11判定出該訓練資料集不包含標籤時，進行子步驟218；當該處理裝置11判定出該訓練資料集包含標籤時，進行子步驟219。

【0025】在子步驟218中，該處理裝置11將該k-平均演算法作為該機器學習演算法。

【0026】在子步驟219中，該處理裝置11根據該訓練資料集判定該訓練資料集之輸入特徵是否為影像型態。當該處理裝置11判定出該訓練資料集之輸入特徵為影像型態時，進行子步驟220；當該處理裝置11判定出該訓練資料集之輸入特徵非為影像型態時，進行子步驟221。

【0027】在子步驟220中，該處理裝置11將該類神經演算法作為該機器學習演算法。

【0028】在子步驟221中，該處理裝置11根據該訓練資料集中的建模資料，分別利用該支持向量機演算法與該類神經演算法進行訓練以分別獲得一第三推論模型與一第四推論模型。

【0029】在子步驟222中，該處理裝置11根據該訓練資料集中的驗證資料，分別利用該第三推論模型與該第四推論模型進行推論以分別獲得一第三推論準確度與一第四推論準確度。

【0030】在子步驟223中，該處理裝置11將具有較高推論準確度之推論模型所對應的演算法作為該機器學習演算法。

【0031】在步驟23中，該處理裝置11根據該訓練資料集，利用該機器學習演算法，進行訓練以獲得一機器學習推論模型。

【0032】在步驟24中，該處理裝置11將訓練完的該機器學習推論模型，利用一轉換套件，轉換成一至少在一可程式化電路裝置12上被運行的執行檔案。

【0033】值得特別說明的是，當該機器學習推論模型屬於一第一類別模型時，其中該第一類別模型為一由該類神經演算法而訓練出的推論模型，該處理裝置11係利用一深度神經網路開發套件(DNNDK)，將訓練完的該機器學習推論模型，轉換成在該可程式化電路裝置12與該處理單元13上被運行的該執行檔案，該執行檔案為一ELF檔，並包含了可運行在該處理單元13上的一第一指令組，及可運行在該可程式化電路裝置12上的一第二指令組，與一暫存器轉換語言(RTL)。該可程式化電路裝置12包括一深度學習處理器(Deep-Learning Processing Unit，簡稱DPU)，該深度學習處理器為一習知的IP core，其利用FPGA內的硬體資源去加速神經網路運算的速度，包括Lookup table、Register、BRAM、DSP組成可調整組態的DPU。此外，當該機器學習推論模型屬於第二類別模型時，其中該第二類別模型為一由該支持向量機演算法或該k-平均演算法而訓練出的推論模型，該處理裝置11係利用一Vivado套件，將訓練完的該機器學習推論模型，轉換成在該可程式化電路裝置12上被運行的執行檔案，該執行檔案為一bitstream檔，該bitstream檔被燒錄到FPGA後，即會形成一個具有根據該支持向量機演算法或該k-平均演算法來進行推論之能力的IP core。

【0034】在步驟25中，該處理單元13將一待分析資料傳送至運行有該執行檔案的該可程式化電路裝置12，以使該可程式化電路裝置

12 進行推論運算以獲得一對應於該待分析資料的推論結果。

【0035】值得一提的是，當該執行檔案係對應於該第一類別模型時，在進行該推論運算時，該處理單元13會透過DPU driver將其指令以及待分析資料傳給DPU做加速運算，DPU運算完再傳回該處理單元13，最後由該處理單元13輸出該推論結果。配合DNNDK所轉換出之分別運行在該處理單元13與該可程式化電路裝置12的第一指令組與第二指令組，及運行在該可程式化電路裝置12上的該暫存器轉換語言以對DPU內平行運算單元進行配置，進而可達成加速神經網路運算的功能。當該執行檔案係對應於該第二類別模型時，在進行該推論運算時，藉由該處理單元13去呼叫FPGA中具有根據該支持向量機演算法或該k-平均演算法來進行推論之能力的IP core，FPGA運算完再傳回該處理單元13，由該處理單元13輸出該推論結果。

【0036】綜上所述，本發明機器學習服務提供方法，藉由該處理裝置11根據該訓練資料集，自動判定出適用該訓練資料集的該機器學習演算法，藉此使用者僅需提供訓練資料集，該處理裝置11即會自動判定出適用該訓練資料集的機器學習演算法，且所判定出之機器學習演算法係較符合該訓練資料集之資料屬性且可獲得具有較高推論準確度的推論模型，進而可大幅降地使用者判定使用何種機器學習方法的負擔，也無需自行設計用於該訓練資料集的機器學習

演算法。接著，藉由該處理裝置11利用該機器學習演算法訓練出該機器學習推論模型，繼而，將訓練完的該機器學習推論模型，轉換成至少在該可程式化電路裝置12上被運行的該執行檔案，藉由該可程式化電路裝置12來進行推論運算，可加速機器學習執行效率。再者，使用具有可重複燒錄以及快速編寫特性的FPGA開發板，讓開發人員能根據需求客製化出可擴展性高的機器學習推論平台，故確實能達成本發明的目的。

【0037】惟以上所述者，僅為本發明的實施例而已，當不能以此限定本發明實施的範圍，凡是依本發明申請專利範圍及專利說明書內容所作的簡單的等效變化與修飾，皆仍屬本發明專利涵蓋的範圍內。

【符號說明】

【0038】

- 1…… 機器學習服務提供系統
- 11…… 處理裝置
- 12…… 可程式化電路裝置
- 13…… 處理單元
- 21~25… 步驟
- 211~223子步驟

【發明申請專利範圍】

【第1項】 一種機器學習服務提供方法，包含以下步驟：

(A)藉由一處理裝置根據訓練資料集，自動判定出一適用該訓練資料集的機器學習演算法，其中，該步驟(A)包含以下子步驟，

(A-1)藉由該處理裝置，根據該訓練資料集判定所欲訓練出之該機器學習推論模型是屬於一預測模型或一分類模型，

(A-2)當該處理裝置判定出所欲訓練出之該機器學習推論模型屬於該預測模型時，藉由該處理裝置根據該訓練資料集判定該訓練資料集之輸入特徵是否為文字型態，及

(A-3)當該處理裝置判定出該訓練資料集之輸入特徵為文字型態時，藉由該處理裝置將一支持向量機演算法作為該機器學習演算法；

(B)藉由該處理裝置根據該訓練資料集，利用該機器學習演算法，進行訓練以獲得一機器學習推論模型；

(C)藉由該處理裝置將訓練完的該機器學習推論模型，利用一轉換套件，轉換成一至少在一可程式化電路裝置上被運行的執行檔案；以及

(D)藉由一電連接該可程式化電路裝置的處理單元，將一待分析資料傳送至運行有該執行檔案的該可程式化電路裝置，以使該可程式化電路裝置進行推論運算以獲得一推論結果。

【第2項】 如請求項1所述的機器學習服務提供方法，其中，在步驟
第1頁，共4頁(發明申請專利範圍)

(A-2)中，該訓練資料集包含建模資料與驗證資料，在該步驟(A-2)後，還包含以下子步驟：

(A-4)當該處理裝置判定出該訓練資料集之輸入特徵非為文字型態時，藉由該處理裝置根據該訓練資料集中的建模資料，分別利用該支持向量機演算法與一類神經演算法進行訓練以分別獲得一第一推論模型與一第二推論模型；

(A-5)藉由該處理裝置根據該訓練資料集中的驗證資料，分別利用該第一推論模型與該第二推論模型進行推論以分別獲得一第一推論準確度與一第二推論準確度；及

(A-6)藉由該處理裝置，將具有較高推論準確度之推論模型所對應的演算法作為該機器學習演算法。

【第3項】 如請求項1所述的機器學習服務提供方法，其中，在步驟(A-1)後，還包含以下子步驟：

(A-7)當該處理裝置判定出所欲訓練出之該機器學習推論模型屬於該分類模型時，藉由該處理裝置根據該訓練資料集判定該訓練資料集是否包含標籤；及

(A-8)當該處理裝置判定出該訓練資料集不包含標籤時，藉由該處理裝置將一k-平均演算法作為該機器學習演算法。

【第4項】 如請求項3所述的機器學習服務提供方法，其中，在步驟(A-7)後，還包含以下子步驟：

(A-9)當該處理裝置判定出該訓練資料集包含標籤時，藉由該處理裝置根據該訓練資料集判定該訓練資料集

之輸入特徵是否為影像型態；及

(A-10)當該處理裝置判定出該訓練資料集之輸入特徵為影像型態時，藉由該處理裝置將一類神經演算法作為該機器學習演算法。

【第5項】 如請求項4所述的機器學習服務提供方法，其中，在步驟(A-2)中，該訓練資料集包含建模資料與驗證資料，在步驟(A-9)後，還包含以下子步驟：

(A-11)當該處理裝置判定出該訓練資料集之輸入特徵非為影像型態時，藉由該處理裝置根據該訓練資料集中的建模資料，分別利用該支持向量機演算法與該類神經演算法進行訓練以分別獲得一第三推論模型與一第四推論模型；

(A-12)藉由該處理裝置根據該訓練資料集中的驗證資料，分別利用該第三推論模型與該第四推論模型進行推論以分別獲得一第三推論準確度與一第四推論準確度；及

(A-13)藉由該處理裝置，將具有較高推論準確度之推論模型所對應的演算法作為該機器學習演算法。

【第6項】 如請求項1所述的機器學習服務提供方法，其中，在步驟(C)中，當該機器學習推論模型屬於一第一類別模型時，該處理裝置係利用一深度神經網路開發套件，將訓練完的該機器學習推論模型，轉換成在該可程式化電路裝置與該處理單元上被運行的執行檔案。

【第7項】 如請求項6所述的機器學習服務提供方法，其中，在步驟(C)中，該第一類別模型為一由一類神經演算法而訓練出

的推論模型。

【第8項】如請求項1所述的機器學習服務提供方法，其中，在步驟(C)中，當該機器學習推論模型屬於第二類別模型時，該處理裝置係利用一Vivado套件，將訓練完的該機器學習推論模型，轉換成在該可程式化電路裝置上被運行的執行檔案。

【第9項】如請求項8所述的機器學習服務提供方法，其中，在步驟(C)中，該第二類別模型為一由一支持向量機演算法及一k-平均演算法之其中一者而訓練出的推論模型。

【發明圖式】

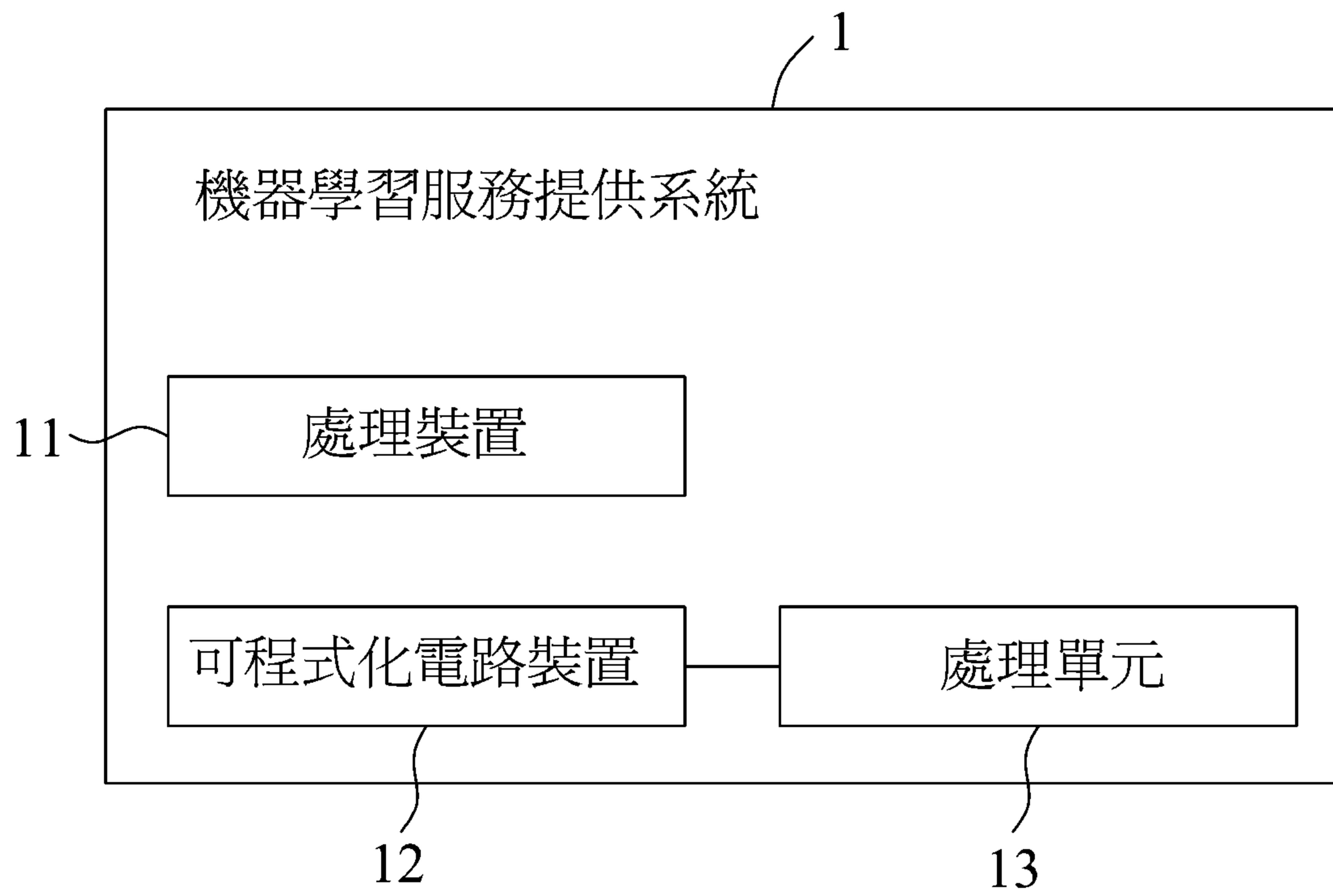


圖1

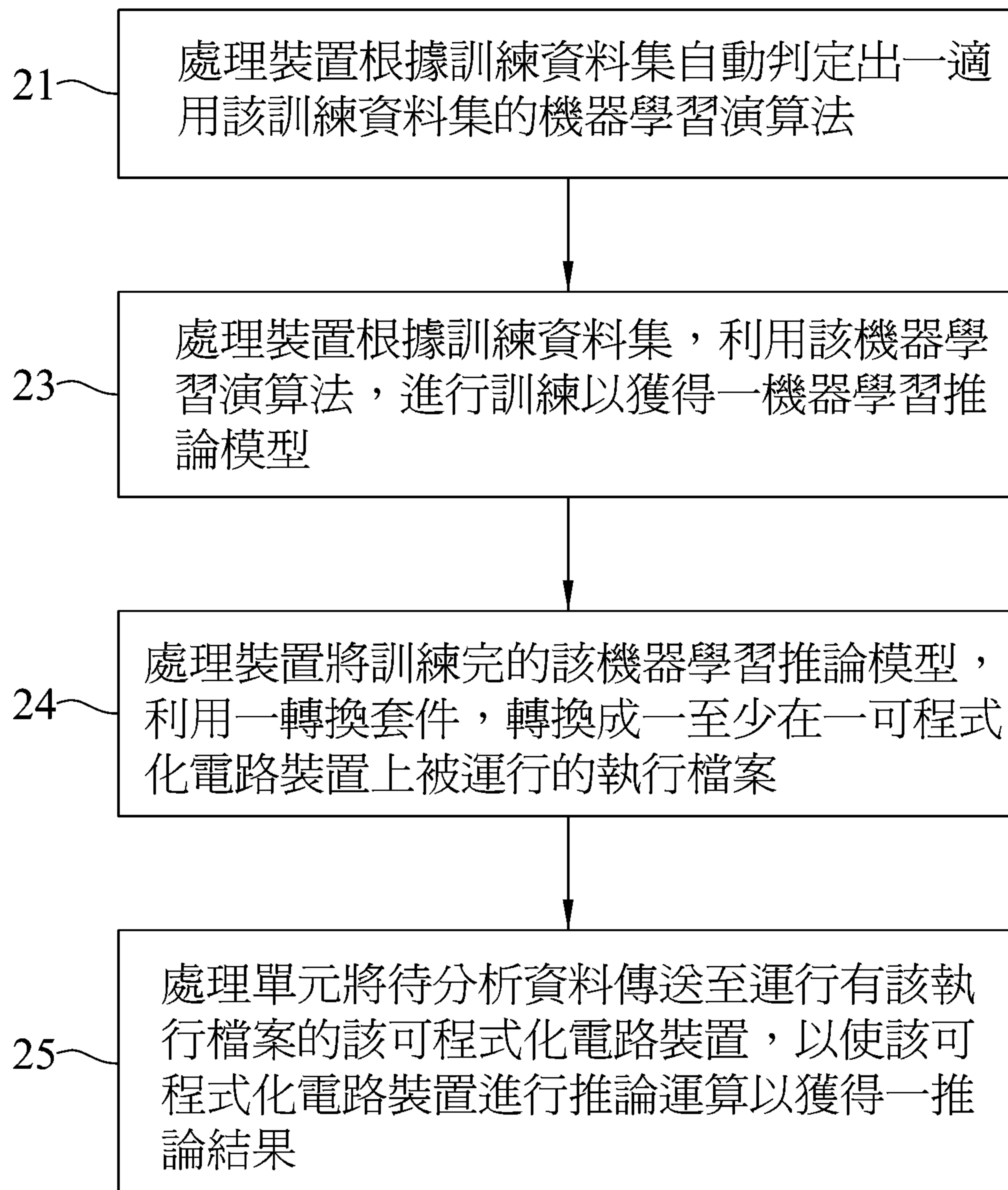


圖2

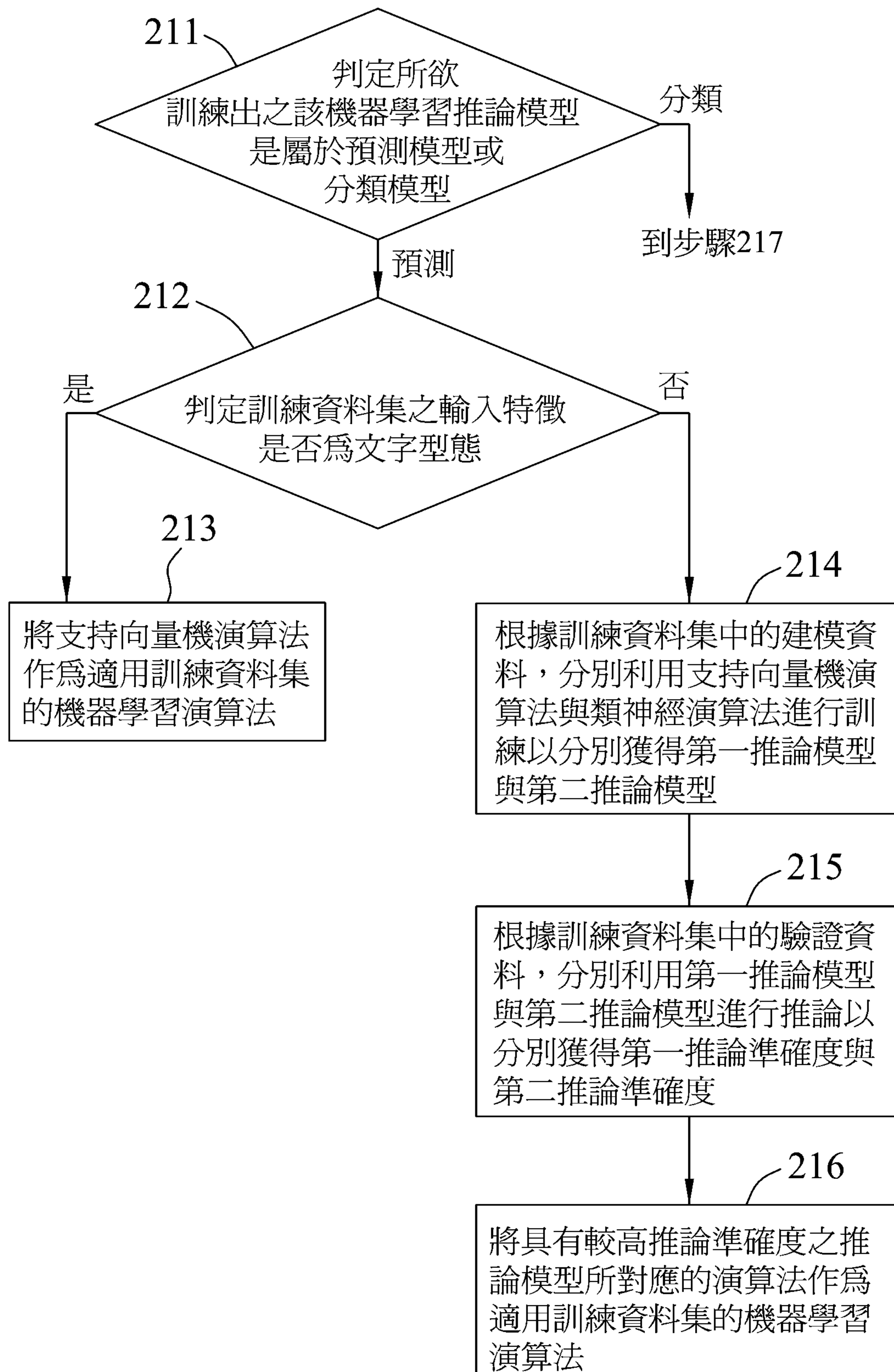


圖3

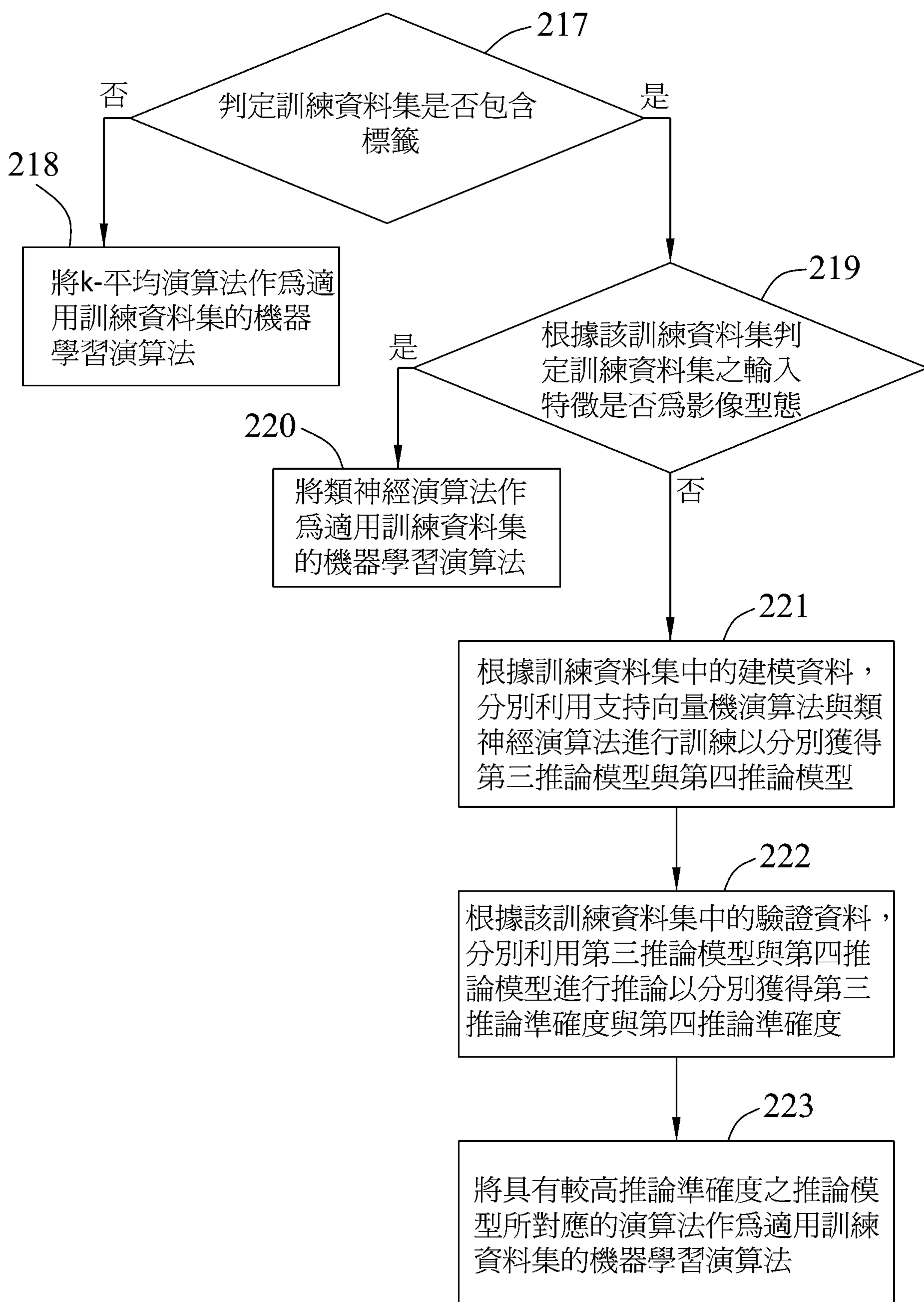


圖4