I767000

(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 I767000 B**

(45)公告日： 中華民國 111 (2022) 年 06 月 11 日

(21)申請案號：107117284 (22)申請日： 中華民國 107 (2018) 年 05 月 21 日

(51)Int. Cl.： ***G06N3/02 (2006.01)***

(30)優先權：2017/05/20 美國 62/509,053

(71)申請人：英商淵慧科技有限公司 (英國) DEEPMIND TECHNOLOGIES LIMITED (GB)
英國

(72)發明人：凡 登 沃爾德 亞倫 傑瑞德 安東尼斯 VAN DEN OORD, AARON GERARD ANTONIUS (BE)；賽門亞 凱倫 SIMONYAN, KAREN (RU)；溫亞爾斯 奧里奧爾 VINYALS, ORIOL (ES)

(74)代理人：陳長文

(56)參考文獻：

網路文獻 Aaron van den Oord, Sander Dieleman, Heiga Zen, Karen Simonyan, Oriol Vinyals, Alex Graves, Nal Kalchbrenner, Andrew Senior, Koray Kavukcuoglu, "WAVENET: A GENERATIVE MODEL FOR RAW AUDIO," September 2016, https://arxiv.org/pdf/1609.03499.pdf WAVENET: A GENERATIVE MODEL FOR RAW AUDIO, WAVENET: A GENERATIVE MODEL FOR RAW AUDIO, WAVENET: A GENERATIVE MODEL FOR RAW AUDIO, WAVENET: A GENERAT

審查人員：吳鴻鎮

申請專利範圍項數：8 項 圖式數：5 共 34 頁

(54)名稱

產生波形之方法及電腦儲存媒體

(57)摘要

一種前饋生成神經網路在一單一神經網路推理中產生包含一特定類型之多個輸出樣本之一輸出實例。視情況，可基於一內容背景輸入來調節該產生。例如，該前饋生成神經網路可產生一語音波形，該語音波形係基於一輸入文字片段之語言特徵調節之該文字片段之一言語表達。

A feedforward generative neural network that generates an output example that includes multiple output samples of a particular type in a single neural network inference. Optionally, the generation may be conditioned on a context input. For example, the feedforward generative neural network may generate a speech waveform that is a verbalization of an input text segment conditioned on linguistic features of the text segment.

指定代表圖：

符號簡單說明：

400:程序

402:步驟

404:步驟

406:步驟

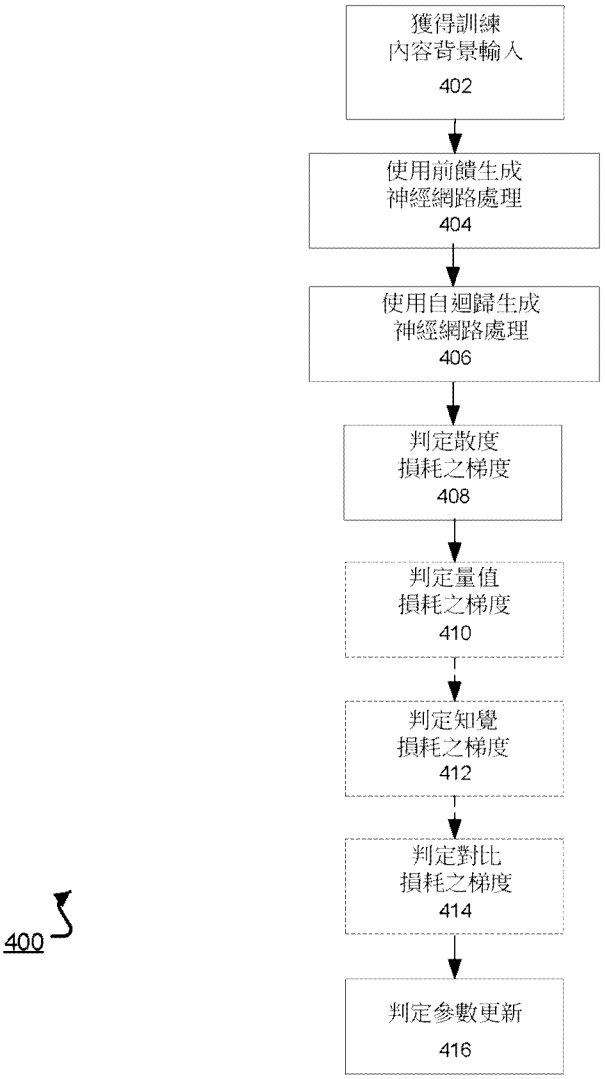
408:步驟

410:步驟

412:步驟

414:步驟

416:步驟



【圖4】

I767000

【發明摘要】

公告本

【中文發明名稱】

產生波形之方法及電腦儲存媒體

【英文發明名稱】

METHOD AND COMPUTER STORAGE MEDIUM OF GENERATING WAVEFORM

【中文】

一種前饋生成神經網路在一單一神經網路推理中產生包含一特定類型之多個輸出樣本之一輸出實例。視情況，可基於一內容背景輸入來調節該產生。例如，該前饋生成神經網路可產生一語音波形，該語音波形係基於一輸入文字片段之語言特徵調節之該文字片段之一言語表達。

【英文】

A feedforward generative neural network that generates an output example that includes multiple output samples of a particular type in a single neural network inference. Optionally, the generation may be conditioned on a context input. For example, the feedforward generative neural network may generate a speech waveform that is a verbalization of an input text segment conditioned on linguistic features of the text segment.

【指定代表圖】

圖4

【代表圖之符號簡單說明】

400　程序

402　步驟

404 步驟

406 步驟

408 步驟

410 步驟

412 步驟

414 步驟

416 步驟

【發明說明書】

【中文發明名稱】

產生波形之方法及電腦儲存媒體

【英文發明名稱】

METHOD AND COMPUTER STORAGE MEDIUM OF GENERATING WAVEFORM

【技術領域】

【先前技術】

【0001】 本說明書係關於生成神經網路。

【0002】 神經網路係採用一或多個非線性單元層以針對一所接收輸入預測一輸出之機器學習模型。除一輸出層以外，一些神經網路亦包含一或多個隱藏層。各隱藏層之輸出被用作網路中之下一層(即，下一隱藏層或輸出層)之輸入。網路之各層根據一各自參數集之當前值從一所接收輸入產生一輸出。

【發明內容】

【0003】 一般言之，本說明書描述一種前饋生成神經網路。一前饋生成神經網路係在一單一神經網路推理中產生包含一特定類型之多個輸出樣本之一輸出實例的一神經網路。視情況，可基於一內容背景輸入調節該產生。例如，該前饋生成神經網路可產生一語音波形，該語音波形係基於一輸入文字片段之語言特徵調節之該文字片段之一言語表達。

【0004】 因此，在一個實施方案中，本說明書描述一種訓練一前饋生成神經網路之方法，該前饋生成神經網路具有複數個前饋參數且經組態以產生基於一第二類型之內容背景輸入調節之一第一類型之輸出實例。各輸出實例包含複數個產生時間步驟之各者處之一各自輸出樣本。該前饋生

成神經網路經組態以接收包括一內容背景輸入之一前饋輸入且處理該前饋輸入以產生一前饋輸出，該前饋輸出針對該等產生時間步驟之各者界定該產生時間步驟處之該輸出樣本之可能值之一各自概率分佈。該訓練包括：獲得一訓練內容背景輸入；使用該前饋生成神經網路根據該等前饋參數之當前值處理包括該訓練內容背景輸入之一訓練前饋輸入以產生一訓練前饋輸出；及使用一受訓練自迴歸生成神經網路處理該訓練內容背景輸入。該受訓練自迴歸生成神經網路已經訓練以針對該複數個產生時間步驟之各者自迴歸地產生一自迴歸輸出，該自迴歸輸出界定基於前述產生時間步驟處之輸出樣本調節之該產生時間步驟處之該輸出樣本之可能值之一概率分佈。該方法進一步包括判定關於該等前饋參數之一第一梯度以最小化一散度損耗。該散度損耗針對該等產生時間步驟之各者取決於從(在)藉由該產生時間步驟之該自迴歸輸出界定之該概率分佈至(與)藉由該訓練前饋輸出界定之該產生時間步驟之該概率分佈(之間)的一第一散度。該方法進一步包括至少部分基於該第一梯度判定對該等前饋參數之該等當前值之一更新。

【0005】 隨後描述此等方法之優點。一般言之但非必要地，該前饋輸入進一步包括該等產生時間步驟之各者處之一各自雜訊輸入。在一些實施方案中，此促進並行產生一組樣本之輸出分佈，因此比一自迴歸手段快得多。

【0006】 該第一散度可為例如一KL散度或一Jensen-Shannon散度。該散度損耗可至少部分取決於該等時間步驟之各者處之該等第一散度之一總和。

【0007】 該訓練可進一步包括獲得該訓練內容背景輸入之一實況輸

出實例且藉由從該等概率分佈取樣而從該訓練前饋輸出產生一預測輸出實例。該預測輸出實例可用於判定該訓練之另一梯度。

【0008】 因此，該實況輸出實例及該預測輸出實例可界定波形，諸如語音波形。該訓練可接著進一步包括：產生該實況輸出實例之一第一量值譜圖；產生該預測輸出實例之一第二量值譜圖；及判定關於該等前饋參數之一第二梯度以最小化取決於該第一量值譜圖與該第二量值譜圖之間的差異之一量值損耗。判定對該等前饋參數之該等當前值之該更新可包括至少部分基於該第二梯度判定該更新。一量值譜圖可包括例如界定一波形之一振幅、能量或類似量值譜圖(例如表示不同頻帶中之功率)之資料。

【0009】 另外或替代地，該訓練可包括使用一受訓練特徵產生神經網路處理該實況輸出實例以獲得該實況輸出實例之特徵。該受訓練特徵產生神經網路可為採用一波形作為輸入之一預訓練神經網路。該訓練可進一步包括使用該受訓練特徵產生神經網路處理該預測輸出實例以獲得該預測輸出實例之特徵。該訓練可接著包括判定關於該等前饋參數之一第三梯度以最小化一知覺損耗。該知覺損耗可界定為取決於該實況輸出實例之該等特徵與該預測輸出實例之該等特徵之間的一差異量度之一損耗。判定對該等前饋參數之該等當前值之該更新可包括至少部分基於該第三梯度判定該更新。

【0010】 特定言之當該等輸出實例包括語音波形時，該特徵產生神經網路可包括一語音辨識神經網路。更一般言之，該特徵產生神經網路可包括一受訓練自迴歸生成神經網路。在此等及其他實施方案中，該等特徵可為該特徵產生網路中之一中間層之輸出。

【0011】 該訓練可另外或替代地包括：獲得一不同內容背景輸入；

使用該受訓練自迴歸生成神經網路處理該不同內容背景輸入以針對該複數個產生時間步驟之各者獲得一各自不同自迴歸輸出；及判定關於該等前饋參數之一第四梯度以最大化一對比損耗。廣義上，此處之對比損耗界定兩個分佈之間的一相似性量度。因此，該對比損耗可界定為針對該等產生時間步驟之各者至少部分取決於來自藉由該產生時間步驟之該不同自迴歸輸出界定之該概率分佈及藉由該訓練前饋輸出界定之該產生時間步驟之該概率分佈之一第二散度之一損耗。判定對該等前饋參數之該等當前值之該更新可接著包括至少部分基於該第四梯度判定該更新。

【0012】 本文亦描述一種產生一輸出實例之方法。該方法可包括接收一內容背景輸入且藉由使用已使用如上文描述之一方法訓練之一前饋生成神經網路處理包括該內容背景輸入之一前饋網路輸入而產生一輸出實例。該前饋網路輸入可包含一雜訊輸入。

【0013】 本文進一步描述一種方法，其包括：接收用以產生一波形之一請求，該波形包括基於表示文字輸入特徵之一調節張量調節之複數個樣本；獲得包含該複數個樣本之各者之一雜訊值之一隨機雜訊向量；及使用一前饋生成神經網路處理該隨機雜訊向量以產生該波形。該前饋生成神經網路可包括卷積神經網路層群組之一序列。可基於該調節張量調節各群組。各群組可經組態以接收一輸入波形且將基於該調節張量調節之該波形變換成一輸出波形。

【0014】 第一卷積神經網路層群組之該輸入波形可為該隨機雜訊向量。除該第一群組以外的各卷積神經網路層群組之該輸入波形可為緊接在該群組之前的群組之一輸出波形。該所產生波形可為最後卷積神經網路層群組之輸出。

【0015】 在一些實施方案中，一卷積神經網路層群組可包括具有藉由該調節張量閘控之一啟動功能之一或多個層。例如，該啟動功能可為使該調節張量與一或多個可學習權重卷積之一功能。各卷積層群組可包含各包含一或多個擴展卷積層之一或多個剩餘區塊，從而促進模型化長期因果相依性。例如，在一些實施方案中，各卷積層群組可具有類似於WaveNet(隨後參考)之一結構，但採用一隨機雜訊向量作為輸入。卷積神經網路(CNN)層群組之序列可包括輸入與輸出之間的CNN層群組之一鏈。

【0016】 該前饋生成神經網路可產生界定該複數個樣本之各者之可能值之一各自概率分佈之一輸出。該處理可進一步包括使用該對應概率分佈選擇該等樣本之各者之一值。

【0017】 本說明書結合系統及電腦程式組件使用術語「組態」。一或多個電腦之一系統經組態以執行特定操作或動作意謂該系統已安裝於其軟體、韌體、硬體或其等之一組合上，軟體、韌體、硬體或其等之一組合經操作以導致該系統執行操作或動作。一或多個電腦程式經組態以執行特定操作或動作意謂一或多個程式包含當藉由資料處理設備執行時導致該設備執行操作或動作之指令。

【0018】 可在特定實施例中實施本說明書中描述之標的物，以實現以下優點之一或多者。

【0019】 自迴歸神經網路藉由在各時間步驟處執行一正向傳送而產生跨多個時間步驟之輸出實例。在一給定時間步驟處，該自迴歸神經網路產生包含於基於已產生之該等輸出樣本調節之該輸出實例中之一新輸出樣本。此可導致一高品質輸出實例但可消耗大量計算資源且花費大量時間，即，因為需要在基於在先前時間步驟處執行之處理調節之較大數目個時間

步驟處執行處理。

【0020】 另一方面，如本說明書中描述之一前饋生成神經網路可比一自迴歸生成神經網路更快地產生輸出實例同時維持該等所產生輸出實例之一高品質程度。

【0021】 特定言之，該所描述前饋生成神經網路在一單一推理步驟中產生該輸出實例，即，透過該神經網路之一單一正向傳送。相對於由一自迴歸神經網路消耗之時間，此極大減少產生該輸出實例所需之時間及計算資源量。

【0022】 另外，由於該前饋神經網路之架構、該神經網路受訓練之方式或兩者，該前饋生成神經網路可產生具有相當於由該受訓練自迴歸神經網路產生之輸出實例之品質之輸出實例。

【0023】 特定言之，該前饋神經網路可接收作為輸入之一雜訊向量且透過(基於該內容背景輸入調節之)多個卷積神經網路層群組處理該雜訊。由於此架構，可有效地基於該內容背景輸入調節該神經網路以產生一高品質輸出實例。另外，該前饋神經網路可具有比該自迴歸神經網路更少之參數及更小之計算複雜性。例如，該自迴歸神經網路亦可為一卷積神經網路，但可為具有比該前饋神經網路更多之參數之一計算上更複雜神經網路。例如，該自迴歸神經網路可包含該前饋神經網路中不存在之跨越連接、具有比該前饋神經網路更大量之隱藏單元或兩者。儘管如此，該前饋神經網路仍可歸因於本說明書中描述之技術而產生具有相稱品質之輸出實例。

【0024】 再者，藉由訓練該前饋神經網路以匹配由一受訓練自迴歸神經網路產生之輸出，該前饋神經網路可經訓練以在一小段時間內產生具

有相當於該受訓練自迴歸神經網路之品質之樣本。可視情況使用其他損耗進一步增強該前饋神經網路之該訓練(例如，一量值損耗、一知覺損耗或一對比損耗之一或多者)以在不增加該前饋神經網路之計算佔用面積的情況下改良該受訓練前饋神經網路之效能。

【0025】 此可允許該前饋神經網路用於在其中產生需要具有低延遲之高品質輸出實例之環境中或在可用於執行該神經網路之計算資源量受限時產生輸出實例。例如，此可在該神經網路經部署於一行動裝置上或具有受限處理功率及記憶體之一專用個人計算裝置(例如，一智慧型揚聲器或其他智慧型裝置)上時發生。

【0026】 在下文之附圖及描述中提出本說明書之標的物之一或多項實施例之細節。自描述、圖式及發明申請專利範圍將變得明白標的物之其他特徵、態樣及優點。

【圖式簡單說明】

【0027】

圖1展示一例示性神經網路系統。

圖2係用於產生一輸出實例之一例示性程序之一流程圖。

圖3係用於訓練一前饋生成神經網路之一例示性程序之一流程圖。

圖4係用於判定對前饋參數之當前值之一更新之一例示性程序之一流程圖。

圖5係用於訓練前饋生成神經網路110之自迴歸生成神經網路130之一圖。

各種圖式中之相同元件符號及命名指示相同元件。

【實施方式】

【0028】 圖1展示一例示性神經網路系統100。神經網路系統100係實施為其中可實施下文描述之系統、組件及技術之一或多個位置中之一或多個電腦上之電腦程式之一系統之一實例。

【0029】 神經網路系統100接收作為輸入之一內容背景輸入102及雜訊104且產生基於內容背景輸入102及雜訊104調節之一輸出實例112。

【0030】 例如，內容背景輸入102可為文字之語言特徵且輸出實例112可為以言語表達(即，說出)之文字之一波形。即，輸出實例可為特性化波形之一值序列，即，振幅值或壓縮或壓縮擴展振幅值之一序列。

【0031】 作為另一實例，內容背景輸入102可為文字且輸出實例112可為由文字描述之一影像。

【0032】 作為又另一實例，內容背景輸入102可為一視訊圖框序列且輸出實例112可為跟隨視訊圖框序列中之最後圖框之一視訊圖框。

【0033】 特定言之，神經網路系統100包含一前饋生成神經網路110。

【0034】 前饋生成神經網路110係經組態以接收內容背景輸入102及雜訊104且產生界定輸出實例112之一前饋輸出之一神經網路。

【0035】 輸出實例112包含多個輸出樣本(即，數值)，且前饋生成神經網路110在一單一推理步驟中(即，在透過前饋生成神經網路110之一單一正向傳送中)產生界定多個數值之各者的前饋輸出，且特定言之未基於藉由神經網路110在任何先前推理步驟處執行的任何處理來進行調節。

【0036】 一般言之，雜訊係包含輸出實例中之樣本之各者之一雜訊值之一隨機雜訊向量。

【0037】 特定言之，前饋輸出針對輸出實例112中之各樣本界定樣

本之可能值之一各自概率分佈。

【0038】 例如，針對各樣本，前饋輸出可包含樣本之一預測值及/或樣本之可能值之一分佈的參數。例如，參數可為樣本之可能值之一邏輯分佈的平均值及標準差。因此，在一些實施方案中，前饋輸出針對各樣本包含樣本之一預測值以及一邏輯分佈的平均值及標準差。如下文描述，在其中CNN層群組經堆疊於序列中之處，僅樣本可從一個層群組傳送至下一層群組，且可藉由各群組從個別平均值及標準差輸出來判定總平均值及標準差。在一些其他實施方案中，前饋輸出可不明確包含一預測樣本值，但可從可能樣本值之分佈的參數判定預測樣本值。例如，預測樣本值可經判定為下列項目的總和：(i)樣本之對應雜訊值與樣本之分佈之平均值的乘積；及(ii)樣本的標準差。

【0039】 當輸出實例係一波形時，前饋生成神經網路110大體上包含卷積神經網路層群組之一序列。基於一調節張量來調節各群組，即，基於內容背景輸入102 (例如，提供語音及持續時間資訊之語言特徵)來進行調節。各群組經組態以接收一輸入波形，且將基於調節張量調節之輸入波形變換成一輸出波形。在https://arxiv.org/pdf/1609.03499.pdf可獲得之WAVENET:A GENERATIVE MODEL FOR RAW AUDIO中描述基於一調節張量調節一卷積層，其之全部內容係以引用的方式併入本文中。因此，對於第一卷積層群組，輸入波形係隨機雜訊向量，且對於最後卷積層群組，輸出係界定輸出實例之前饋輸出。

【0040】 在下文參考圖5更詳細描述前饋生成神經網路110之例示性架構。在下文參考圖2更詳細描述使用前饋生成神經網路110產生一輸出實例。

【0041】 為了訓練前饋生成神經網路110以產生精確前饋輸出，神經網路系統100亦包含一訓練子系統120，其訓練前饋生成神經網路110以判定前饋生成神經網路110之參數(在本說明書中稱為「前饋參數」)之受訓練值。

【0042】 一旦已訓練前饋生成神經網路110，網路110便可經部署且用於針對新接收之內容背景輸入產生輸出實例。例如，網路110可在具有受限計算資源或需要憑藉極低延遲產生語音之一使用者裝置(例如，一行動裝置或一專用智慧型裝置)上實施。

【0043】 特定言之，訓練子系統120使用一受訓練自迴歸生成神經網路130訓練前饋生成神經網路110。

【0044】 自迴歸生成神經網路130亦經組態以接收與前饋生成神經網路110相同類型之內容背景輸入且產生與由前饋生成神經網路110產生之輸出實例相同類型之輸出實例。然而，自迴歸生成神經網路130經組態以在多個時間步驟內以一自迴歸方式產生一輸出實例中之值，即，產生基於在先前時間步驟處產生之輸出實例中之先前樣本之值調節之輸出實例中之各樣本。

【0045】 例如，當輸入係語言特徵且輸出係波形時，自迴歸生成神經網路130可為一自迴歸卷積神經網路。在https://arxiv.org/pdf/1609.03499.pdf可獲得之WAVENET:A GENERATIVE MODEL FOR RAW AUDIO中更詳細描述此一自迴歸神經網路之一實例。

【0046】 一般言之，自迴歸生成神經網路130一旦經訓練，便能夠產生非常高品質輸出。然而，由於前饋神經網路110在一單一傳送中產生

輸出實例，所以前饋神經網路110可能夠產生具有比自迴歸神經網路130低得多的延遲之輸出實例。

【0047】 如將在下文更詳細描述，當使用自迴歸生成神經網路130訓練前饋神經網路110時，訓練子系統使自迴歸神經網路130之參數之值保持固定且使用由自迴歸神經網路130產生之輸出以評估由前饋神經網路110在訓練期間產生之輸出之品質。

【0048】 在下文參考圖5更詳細描述自迴歸生成神經網路130之例示性架構。

【0049】 視情況，訓練子系統120亦使用一特徵產生神經網路140訓練前饋生成神經網路110。

【0050】 特徵產生神經網路140係經組態以接收與由前饋生成神經網路110及自迴歸生成神經網路130產生之輸出實例相同類型之輸入且產生輸入之特徵作為處理輸入之部分之一神經網路。特徵產生神經網路140大體上可為處理輸入以基於輸入產生一得分、分類或迴歸輸出之任何神經網路。

【0051】 特定言之，訓練子系統120使用由特徵產生神經網路140產生之特徵來訓練前饋生成神經網路110。特徵可為網路140之輸出層之輸出、網路140之一中間層之輸出或網路140之兩個或兩個以上層之輸出之一組合。

【0052】 例如，特徵產生神經網路140可為將與輸出實例相同類型之一輸入轉換成與內容背景輸入相同類型之一輸出之一神經網路且特徵可為特徵產生神經網路140之隱藏層之一或多者之輸出。

【0053】 即，當輸出實例係波形且內容背景資料係文字時，特徵產

生神經網路140可為將一話語之波形轉換成話語之一轉錄之一語音辨識神經網路。

【0054】 作為另一實例，特徵產生神經網路140及自迴歸生成神經網路130可為相同神經網路且特徵可為自迴歸生成神經網路130之隱藏層之一或多者之輸出，即，並非概率分佈係自迴歸生成神經網路130之輸出層之輸出。

【0055】 特徵產生神經網路140及自迴歸生成神經網路130兩者在用於訓練前饋神經網路110之前被充分訓練。

【0056】 在下文參考圖3及圖4更詳細描述訓練前饋生成神經網路110。

【0057】 圖2係用於使用前饋生成神經網路產生一輸出實例之一例示性程序200之一流程圖。為方便起見，程序200將描述為由定位於一或多個位置中之一或多個電腦之一系統執行。例如，適當程式化之一神經網路系統(例如，圖1之神經網路系統100)可執行程序200。

【0058】 系統接收呈一調節張量之形式之一內容背景輸入(步驟202)。

【0059】 系統獲得用於產生輸出實例之雜訊(步驟204)。特定言之，雜訊係包含輸出實例中之樣本之各者之一雜訊值之一隨機雜訊向量。例如，系統可對來自一預定分佈(例如，一邏輯分佈)之雜訊向量中之各雜訊值取樣。

【0060】 系統使用前饋生成神經網路處理雜訊向量以產生輸出實例(步驟206)。即，系統使用前饋生成神經網路處理雜訊向量，同時基於調節張量調節神經網路。當前饋生成神經網路包含多個卷積層群組之一序列

時，各群組接收一輸入波形且將輸入波形映射至界定一輸出波形之一輸出。對於第一群組，輸入波形係雜訊向量且對於各其他群組，輸入波形係藉由序列中之前述群組之輸出界定之波形。

【0061】 圖3係用於訓練一前饋生成神經網路之一例示性程序300之一流程圖。為方便起見，程序300將描述為由定位於一或多個位置中之一或多個電腦之一系統執行。例如，適當程式化之一神經網路系統(例如，圖1之神經網路系統100)可執行程序300。

【0062】 系統獲得規定一受訓練自迴歸生成神經網路之資料(步驟302)。

【0063】 視情況，系統獲得規定一特徵產生神經網路之資料(步驟304)。如上文描述，特徵產生神經網路可為與自迴歸神經網路相同或不同之一網路。

【0064】 系統使用受訓練自迴歸生成神經網路及視情況受訓練特徵產生神經網路訓練一前饋生成神經網路(步驟306)。

【0065】 特定言之，系統訓練前饋生成模型以藉由最佳化至少部分取決於由前饋生成模型產生之概率分佈與由受訓練自迴歸生成神經網路產生之概率分佈之間的一散度之一目標函數而從前饋參數之初始值判定前饋參數之受訓練值。

【0066】 在訓練期間，系統調整前饋參數之值，同時使自迴歸參數之受訓練值及(若使用)特徵產生參數之受訓練值保持固定。

【0067】 特定言之，系統訓練前饋生成神經網路以最小化包含一散度損耗及視情況一量值損耗、一知覺損耗或一對比損耗之一或多者之一損耗函數。當損耗函數包含多個損耗時，損耗函數可為損耗之一加權總和。

將在下文參考圖4更詳細描述個別損耗。

【0068】 圖4係用於判定對前饋參數之當前值之一更新之一例示性程序400之一流程圖。為方便起見，程序400將描述為由定位於一或多個位置中之一或多個電腦之一系統執行。例如，適當程式化之一神經網路系統(例如，圖1之神經網路系統100)可執行程序400。

【0069】 系統可重複執行程序400以藉由重複調整前饋參數之值而訓練前饋生成神經網路。

【0070】 系統獲得一訓練內容背景輸入(步驟402)。例如，系統可對一批訓練內容背景輸入取樣且所獲得訓練內容背景輸入可為該批中之輸入之一者。

【0071】 系統使用前饋生成神經網路根據前饋參數之當前值處理包含訓練內容背景輸入之一訓練前饋輸入以產生界定一輸出實例之一訓練前饋輸出(步驟404)。如上文描述，訓練前饋輸出包含產生時間步驟之各者之一各自概率分佈之參數。

【0072】 系統使用受訓練自迴歸生成神經網路處理訓練內容背景輸入以針對複數個產生時間步驟之各者產生一各自自迴歸輸出(步驟406)。如上文描述，一給定產生時間步驟之自迴歸輸出亦界定該時間步驟之一概率分佈。在各時間步驟處，系統基於由前饋神經網路產生之輸出實例之對應部分調節自迴歸生成神經網路，即，使得受訓練自迴歸生成神經網路用於為由前饋神經網路產生之輸出實例打分。即，當產生對應於輸出實例中之一給定樣本之自迴歸輸出時，基於訓練內容背景輸入及由前饋神經網路產生之在輸出實例中之給定樣本之前的樣本調節自迴歸生成神經網路。此在圖5中以圖形描述且在隨附描述中更詳細描述。

【0073】 系統判定關於前饋參數之一散度損耗之一梯度(即，用以最小化散度損耗之一梯度)。散度損耗針對產生時間步驟之各者取決於來自藉由產生時間步驟之自迴歸輸出界定之概率分佈及藉由訓練前饋輸出界定之產生時間步驟之概率分佈之一散度(步驟408)。特定言之，散度損耗可為產生時間步驟之各者之散度之一總和。系統可使用量測一個概率分佈如何從另一概率分佈發散之各種散度量度之任一者。例如，散度可為一KL(Kullback-Leibler)散度。作為另一實例，散度可為一Jensen-Shannon散度。

【0074】 藉由最小化此散度損耗，系統訓練前饋網路以嘗試在已藉由生成神經網路學習之分佈下匹配其自身樣本之概率。

【0075】 在其中採用KL散度D_{KL}之一情況中，此可藉由從前饋分佈與自迴歸分佈之一交叉熵H_{FA}減去前饋模型分佈之一熵H_F ($D_{KL}=H_{FA}-H_F$)而計算，其中H_F給定為：

$$\sum_{t=1}^{T}\left[\sum_{t=1}^{T}\ln s(z_{<t},\theta)\right]+2T$$

【0076】 對於T個樣本，其中$s(z_{<t},\theta)$係(基於輸出實例中在t之前的樣本之雜訊值z且根據前饋參數θ之當前值產生之)樣本t之前饋模型之標準差且T係在該輸出實例中之樣本的總數。交叉熵項H_{FA}給定為：

$$\sum_{t=1}^{T}\mathbb{E}_{p_F(x_{<t})}H\left(p_F(x_t|x_{<t}),p_A(x_t|x_{<t})\right)$$

【0077】 此處，對於從前饋模型分佈p_F提取之每一樣本x_t，可並行判定來自自迴歸模型之值p_A且接著可藉由從各時間步驟t之p_F提取多個不同樣本而評估熵項$H(p_F, p_A)$。

【0078】 當輸出實例係波形時，系統視情況亦判定關於前饋參數之

一量值損耗(亦稱為功率損耗)之一梯度，即，用以最小化關於前饋參數之量值損耗之一梯度(步驟410)。

【0079】 特定言之，為了判定此梯度，系統獲得訓練內容背景輸入之一實況輸出實例。實況輸出實例係應藉由自迴歸及前饋生成神經網路針對訓練內容背景輸入產生之輸出實例，即，訓練內容背景輸入之一已知輸出。系統接著產生實況訓練實例之一量值譜圖。

【0080】 系統亦從訓練前饋輸出產生一預測輸出實例，即，藉由對來自產生時間步驟之各者處之概率分佈之一輸出樣本取樣。系統亦產生預測輸出實例之一量值譜圖。

【0081】 系統接著判定關於取決於實況輸出實例之量值譜圖與預測輸出實例之量值譜圖之間的差異之一量值損耗之前饋參數之一梯度，即，用以鼓勵兩個輸出實例具有類似量值譜圖。此可為例如實況波形及預測波形之短期傅立葉變換(Short Term Fourier Transform)之一方差。包含此一量值損耗項可幫助減小前饋生成模型崩潰至類似低語之一高熵模式之風險。

【0082】 系統進一步視情況判定關於訓練內容背景輸入之前饋參數之一知覺損耗之一梯度(步驟412)。

【0083】 為判定此梯度，系統使用受訓練特徵產生神經網路處理實況輸出實例以獲得實況輸出實例之特徵且接著使用受訓練特徵產生神經網路處理預測輸出實例以獲得預測輸出實例之特徵。

【0084】 系統接著判定用以最小化取決於實況輸出實例之特徵與預測輸出實例之特徵之間的一差異量度之一知覺損耗之一梯度。例如，當特徵係具有尺寸$L_k \times C_k$之一層之輸出時，知覺損耗$loss_{aux}$可滿足：

$$G_{ij} = \frac{1}{L_k C_k} o_k(x)^T o_k(x)$$

$$loss_{aux} = ||G_{ij}(x) - G_{ij}(\hat{x})||_2^2$$

其中$\varphi_k(x)$係實況輸出實例之特徵且$\varphi_k(\hat{x})$係預測輸出實例之特徵。包含此一知覺損耗項可幫助確保良好特徵表示，例如特徵可包含辨識電話用以處罰較差發音之特徵。

【0085】 作為另一實例，損耗可為特徵之間的歐幾里德距離(Euclidean distance)。

【0086】 系統進一步視情況判定關於前饋參數之一對比損耗之一梯度(步驟414)。

【0087】 為判定此梯度，系統獲得一不同內容背景輸入，即，一內容背景輸入之實況輸出實例係不同於訓練內容背景輸入之實況輸出實例。

【0088】 系統使用受訓練自迴歸生成神經網路處理不同內容背景輸入，以針對複數個產生時間步驟之各者獲得一各自不同自迴歸輸出。即，系統基於不同內容背景輸入來調節受訓練自迴歸生成神經網路，同時亦繼續基於由前饋神經網路產生之輸出樣本來調節受訓練自迴歸生成神經網路。

【0089】 系統接著判定關於前饋參數之一梯度，以最大化針對產生時間步驟之各者之至少部分取決於從藉由產生時間步驟之不同自迴歸輸出界定之概率分佈至藉由訓練前饋輸出界定之產生時間步驟之概率分佈之一散度之一對比損耗(或最小化對比損耗之負值)。例如，對比損耗可為調節向量相同時之散度損耗與調節向量不同時之損耗之間之一差異，視情況具有此等項之一相對加權。如同散度損耗，對比損耗可為產生時間步驟之各者之個別散度之一總和，且散度量度可相同於用於散度損耗之散度量度。

因此，除在基於相同資訊來調節前饋模型及自迴歸模型時最小化散度損耗以外，對比損耗亦可在基於不同資訊調節模型時最大化散度損耗。此不利於具有一高概率之波形而無關於調節向量。

【0090】 系統判定對前饋參數之當前值之一更新(步驟416)。特定言之，系統從散度損耗之梯度及使用時之量值損耗、知覺損耗及對比損耗之梯度來判定更新。如上文描述，當使用多個損耗時，更新係該等損耗之一加權總和。

【0091】 系統大體上針對一批訓練內容背景輸入中之各訓練內容背景輸入重複程序400以產生對各訓練內容背景輸入之當前值之一各自更新且接著使更新相加以產生一最終更新。系統可接著使用最終更新更新參數之當前值。系統更新當前值之方式取決於系統在更新參數時所使用之最佳化器。例如，當使用隨機梯度下降時，系統可將一學習率應用至最終更新且接著將結果加減至參數之當前值。

【0092】 圖5係用於訓練前饋生成神經網路110之自迴歸生成神經網路130之一圖500。如圖5中展示，兩個神經網路經組態以產生波形且兩個神經網路之調節輸入102係一文字片段之語言特徵。

【0093】 另外，如圖5中展示，兩個神經網路係卷積神經網路。特定言之，圖5之簡化實例展示具有一單一擴展卷積層群組之兩個神經網路，其中擴展在群組中之各層之後增大。然而，實際上，兩個神經網路可包含多個擴展卷積層群組之一序列，其中各群組經組態以接收一輸入波形且將基於調節張量(即，基於語言特徵)調節之波形變換成一輸出波形。特定言之，兩個神經網路中之卷積層具有不僅基於藉由該層執行之卷積之輸出調節而且亦基於調節張量調節之閘控啟動功能，如在

https://arxiv.org/pdf/1609.03499.pdf 可獲得之WAVENET:A GENERATIVE MODEL FOR RAW AUDIO中描述。即，擴展可在各群組中之最後一層之後重設。在一些情況中，兩個網路亦包含從一給定群組之輸入至給定群組之輸出之剩餘連接。

【0094】 一般言之，兩個神經網路之架構可為較大程度上相同的。例如，兩個網路可包含相同數目個卷積層群組。然而，前饋生成神經網路110經組態以接收雜訊輸入104，而自迴歸生成神經網路130需要基於一部分產生之輸出實例(圖5之「所產生樣本」)進行調節。因此，前饋生成神經網路110可在一單一推理步驟內產生輸出實例，而自迴歸生成神經網路130需要許多推理步驟來產生一輸出實例。

【0095】 另外，在一些實施方案中，前饋生成神經網路110之架構在計算上比自迴歸生成神經網路130之架構較不複雜。例如，雖然兩個網路具有相同數目個卷積層群組，但自迴歸生成神經網路130可具有連接各卷積層群組之輸出之跨越連接，而前饋生成神經網路110不具有跨越連接。作為另一實例，前饋生成神經網路110可在網路之閘控及剩餘組件中具有比自迴歸生成神經網路130更少之隱藏單元。

【0096】 現在參考圖5之實例，在前饋網路110之訓練期間的一給定反覆處，前饋網路110已產生界定包含所產生樣本x_0至x_i之一輸出實例之一前饋輸出502。例如，如上文描述，前饋輸出502可針對各輸出樣本包含輸出樣本之一所產生值及輸出樣本之可能值之一分佈之參數。作為另一實例，亦如上文描述，前饋輸出502可僅包含分佈之參數且可基於對應雜訊值及參數判定樣本。

【0097】 接著在一給定產生時間步驟處將由前饋網路110產生之輸

出實例中之所產生樣本x_0至x_{i-1} (即，x_i之前的所有樣本)作為輸入饋送至自迴歸網路130。在給定產生時間步驟處，自迴歸網路130產生界定輸出實例中之輸出樣本i之可能值之一概率分佈之一自迴歸輸出130。

【0098】 兩個概率分佈(即，由前饋神經網路110產生之概率分佈及由基於由前饋神經網路110產生之輸出樣本調節之自迴歸神經網路130產生之概率分佈)用於如上文描述般訓練前饋神經網路110而不調整受訓練自迴歸神經網路130。因此，受訓練自迴歸神經網路130用於評估由前饋神經網路110產生之輸出之品質。

【0099】 本說明書中描述之標的物及功能操作之實施例可依以下各者實施：數位電子電路、有形體現之電腦軟體或韌體、電腦硬體(包含本說明書中揭示之結構及其等結構等效物)，或其等之一或多者之組合。本說明書中描述之標的物之實施例可實施為一或多個電腦程式，即，在一有形非暫時性程式載體上編碼以供資料處理設備執行或控制該資料處理設備之操作之電腦程式指令之一或多個模組。替代地或另外，程式指令可在一人工產生的傳播信號(例如，一機器產生電、光學或電磁信號)上編碼，產生該信號以編碼資訊以供傳輸至適當接收器設備從而供一資料處理設備執行。電腦儲存媒體可為一機器可讀儲存裝置、一機器可讀儲存基板、一隨機或串列存取記憶體裝置或其等之一或多者之一組合。

【0100】 術語「資料處理設備」涵蓋用於處理資料之各種設備、裝置及機器，包含例如一可程式化處理器、一電腦或多個處理器或電腦。設備可包含專用邏輯電路，例如一FPGA(場可程式化閘極陣列)或一ASIC(特定應用積體電路)。設備除包含硬體之外，亦可包含產生所述電腦程式之一執行環境之程式碼，例如構成處理器韌體、一協定堆疊、一資

料庫管理系統、一作業系統或其等之一或多者之一組合之程式碼。

【0101】 可用任何形式之程式設計語言撰寫一電腦程式(其亦可被稱為或描述為一程式、軟體、一軟體應用程式、一應用程式、一模組、一軟體模組、一指令檔或程式碼)，包含編譯或解譯語言或宣告或程序語言；且其可用任何形式予以部署，包含作為一獨立程式或作為一模組、組件、副常式或適用於一運算環境中之其他單元。一電腦程式可(但無需)對應於一檔案系統中之一檔案。一程式可儲存於保存其他程式或資料(例如儲存於一標記語言文件中之一或多個指令檔)之一檔案之一部分中、儲存於專用於所述程式之一單一檔案中或儲存於多個協同檔案(例如儲存一或多個模組、副程式或程式碼之部分之檔案)中。一電腦程式可經部署以在一個電腦上或在定位於一個位置處或跨多個位置分佈且由一通信網路互連之多個電腦上執行。

【0102】 本說明書中描述之程序及邏輯流程可由一或多個可程式化電腦執行，該一或多個可程式化電腦執行一或多個電腦程式以藉由對輸入資料進行操作及產生輸出而執行功能。亦可藉由專用邏輯電路(例如一FPGA (場可程式化閘極陣列)或一ASIC (特定應用積體電路))執行程序及邏輯流程，且設備亦可實施為該專用邏輯電路。

【0103】 藉由實例，適用於執行一電腦程式之電腦包含、可基於通用或專用微處理器或兩者，或任何其他種類之中央處理單元。一般言之，一中央處理單元將從一唯讀記憶體或一隨機存取記憶體或兩者接收指令及資料。一電腦之基本元件係用於執行(perform或execute)指令之一中央處理單元及用於儲存指令及資料之一或多個記憶體裝置。一般言之，一電腦亦將包含用於儲存資料之一或多個大容量儲存裝置(例如，磁碟、磁光碟

或光碟)，或可操作地耦合以自該大容量儲存裝置接收資料或傳送資料至該大容量儲存裝置，或既自該大容量儲存裝置接收資料亦傳送資料至該大容量儲存裝置。然而，一電腦未必具有此等裝置。而且，一電腦可嵌入另一裝置中，例如一行動電話、一個人數位助理(PDA)、一行動音訊或視訊播放機、一遊戲控制台、一全球定位系統(GPS)接收器或一可攜式儲存裝置(例如一通用串列匯流排(USB)快閃隨身碟)(此處僅列舉一些)。適用於儲存電腦程式指令及資料之電腦可讀媒體包含所有形式之非揮發性記憶體、媒體及記憶體裝置，包含例如半導體記憶體裝置，例如EPROM、EEPROM及快閃記憶體裝置；磁碟，例如內部硬碟或可卸除式磁碟；磁光碟；及CD ROM及DVD-ROM磁碟。處理器及記憶體可藉由專用邏輯電路補充或併入專用邏輯電路中。

【0104】 為提供與一使用者之互動，本說明書中描述之標的物之實施例可在具有用於顯示資訊給使用者之一顯示裝置(例如一CRT (陰極射線管)或LCD (液晶顯示器)監視器)及一鍵盤及一指標裝置(例如一滑鼠或一軌跡球，使用者可藉由其等提供輸入至電腦)之一電腦上實施。其他類型之裝置亦可用於提供與一使用者之互動；例如，提供至使用者之回饋可為任何形式之感測回饋，例如視覺回饋、聽覺回饋或觸覺回饋；且來自使用者之輸入可經接收為任何形式，包含聲學、語音或觸覺輸入。另外，一電腦可藉由將文件發送至供一使用者使用之一裝置及自該裝置接收文件(例如藉由回應於自一使用者之用戶端裝置上之一網頁瀏覽器接收之請求而將網頁發送至網頁瀏覽器)而與使用者互動。

【0105】 可在一計算系統中實施本說明書中描述之標的物之實施例，該計算系統包含一後端組件(例如作為一資料伺服器)，或包含一中介

軟體組件(例如一應用程式伺服器)，或包含一前端組件(例如一用戶端電腦，其具有一圖形使用者介面、一網頁瀏覽器或一應用程式，一使用者可透過其等與本說明書中描述之標的物之一實施方案互動)或一或多個此後端組件、中介軟體組件或前端組件之任何組合。系統之組件可藉由數位資料通信之任何形式或媒體(例如一通信網路)互連。通信網路之實例包含一區域網路(「LAN」)及一廣域網路(「WAN」) (例如網際網路)。

【0106】 計算系統可包含用戶端及伺服器。一用戶端及伺服器通常彼此遠離且通常透過一通信網路而互動。用戶端與伺服器之關係憑藉在各自電腦上運行且彼此具有一用戶端-伺服器關係之電腦程式而引起。

【0107】 雖然本說明書含有許多特定實施方案細節，但此等不應被解釋為對本發明之範疇或者可主張之內容之限制，而應當解釋為可特定於本發明之特定實施例之特徵之描述。本說明書中在單獨實施例之內容脈絡中描述之某些特徵亦可在一單一實施例中組合實施。相反地，在一單一實施例之內容脈絡中描述之各種特徵亦可以單獨地或者以任何合適之子組合在多個實施例中實施。再者，儘管上文可將特徵描述為以特定組合起作用且即使最初如此主張，但在一些情況中，來自所主張組合之一或多個特徵可自組合中切除，且所主張組合可關於一子組合或一子組合之變動。

【0108】 類似地，雖然在圖式中按一特定順序描繪操作，但此不應被理解為要求按所展示之特定順序或循序順序執行此等操作，或執行所有繪示之操作以達成所要結果。在特定境況中，多任務處理及平行處理可為有利的。而且，在上文中描述之實施例中之各種系統模組及組件之分離不應被理解為在所有實施例中皆需要此分離，且應理解所描述之程式組件及系統可大體上一起整合於一單一軟體產品中或封裝至多個軟體產品中。

【0109】已描述標的物之特定實施例。其他實施例在下列發明申請專利範圍之範疇內。例如，在發明申請專利範圍中敘述之動作可按一不同順序執行且仍達成所要結果。作為一個實例，在附圖中描繪之程序不必要求所展示之特定順序或連續順序來達成所要結果。在特定實施方案中，多任務處理及平行處理可為有利的。

【符號說明】

【0110】

100 神經網路系統

102 內容背景輸入

104 雜訊

110 前饋生成神經網路

112 輸出實例

120 訓練子系統

130 自迴歸生成神經網路

140 特徵產生神經網路

200 程序

202 步驟

204 步驟

206 步驟

300 程序

302 步驟

304 步驟

306 步驟

400 程序

402 步驟

404 步驟

406 步驟

408 步驟

410 步驟

412 步驟

414 步驟

416 步驟

500 圖

502 前饋輸出

【發明申請專利範圍】

【第1項】

一種產生波形之方法，其包括：

接收用以產生一波形之一請求，該波形包括基於表示文字輸入特徵之一調節張量來調節之複數個樣本；

獲得包含該複數個樣本之各者之一雜訊值之一隨機雜訊向量；

使用一前饋生成神經網路處理該隨機雜訊向量以產生該波形，其中該前饋生成神經網路包括卷積神經網路層群組之一序列，其中基於該調節張量來調節各群組，且其中各群組經組態以接收一輸入波形，且將基於該調節張量所調節之該波形變換成一輸出波形。

【第2項】

如請求項1之方法，其中該所產生波形係最後卷積神經網路層群組之輸出。

【第3項】

如請求項1或2之方法，其中該第一卷積神經網路層群組之該輸入波形係該隨機雜訊向量。

【第4項】

如請求項1或2之方法，其中除該第一群組以外之各卷積神經網路層群組之該輸入波形係緊接在該群組之前的群組之一輸出波形。

【第5項】

如請求項1或2之方法，其中該前饋生成神經網路產生界定該複數個樣本之各者之可能值之一各自概率分佈之一輸出，且其中該處理進一步包括使用該對應概率分佈來選擇該等樣本之各者之一值。

【第6項】

如請求項1或2之方法，其中各卷積層群組包含各包含一或多個擴展卷積層中的一或多個剩餘區塊。

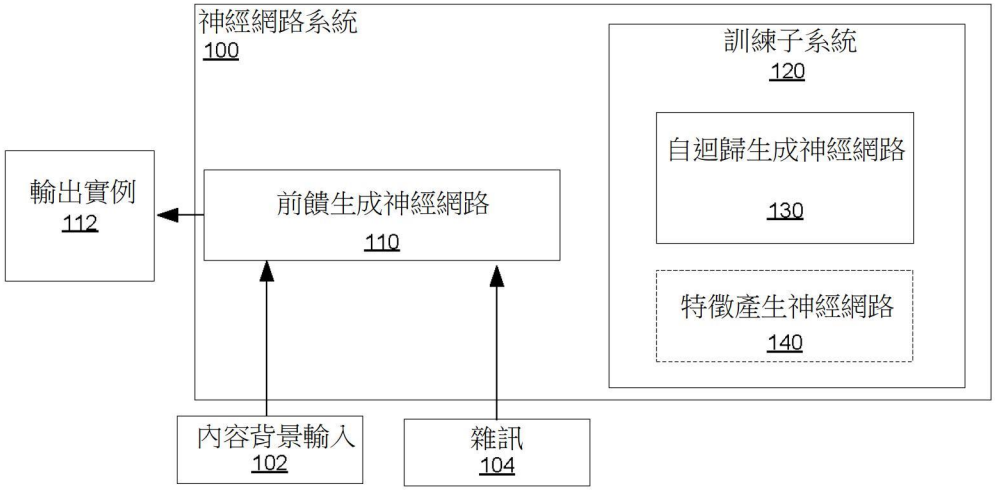
【第7項】

一種儲存有訓練一前饋生成神經網路之指令之電腦儲存媒體，該等指令當藉由一或多個電腦實施時導致該一或多個電腦執行如請求項1至6中任一項之各自方法之操作。

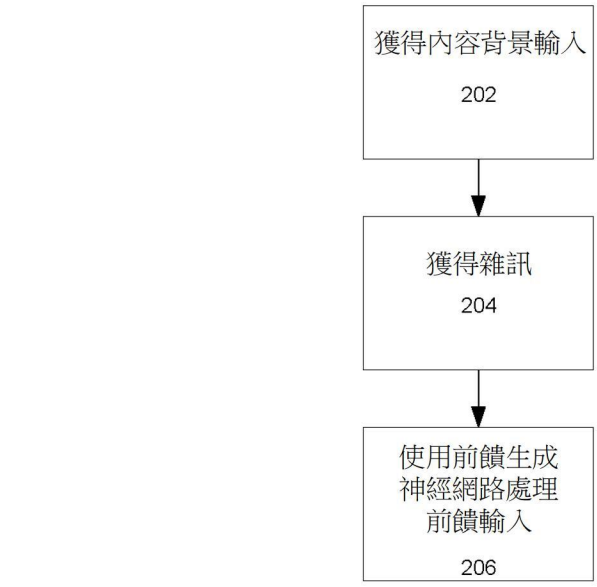
【第8項】

一種包含一或多個電腦及一或多個儲存裝置之系統，該等儲存裝置儲存指令當藉由該一或多個電腦實施時導致該一或多個電腦執行如請求項1至6中任一項之各自方法之操作。

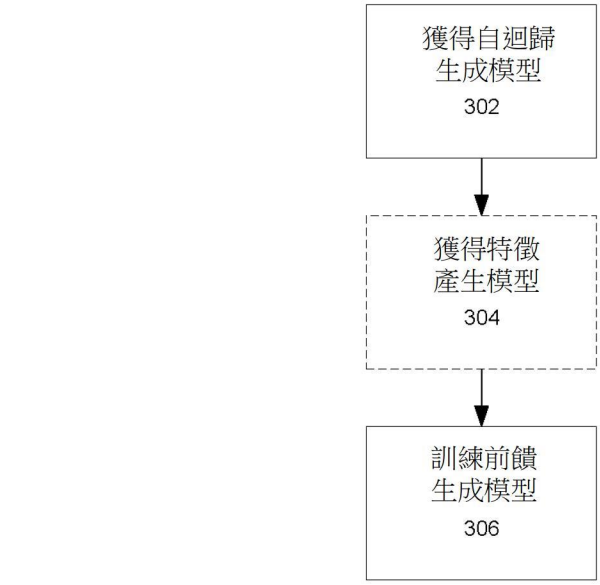
【發明圖式】


神經網路系統
100
訓練子系統
120
自迴歸生成神經網路
130
特徵產生神經網路
140
輸出實例
112
前饋生成神經網路
110
內容背景輸入
102
雜訊
104

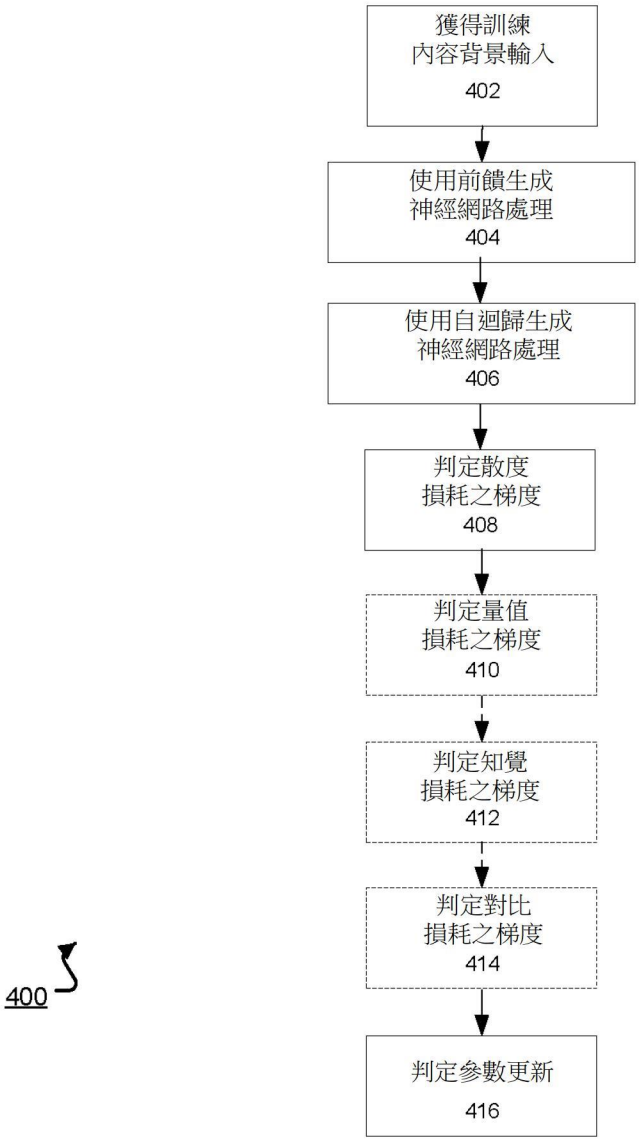
【圖1】



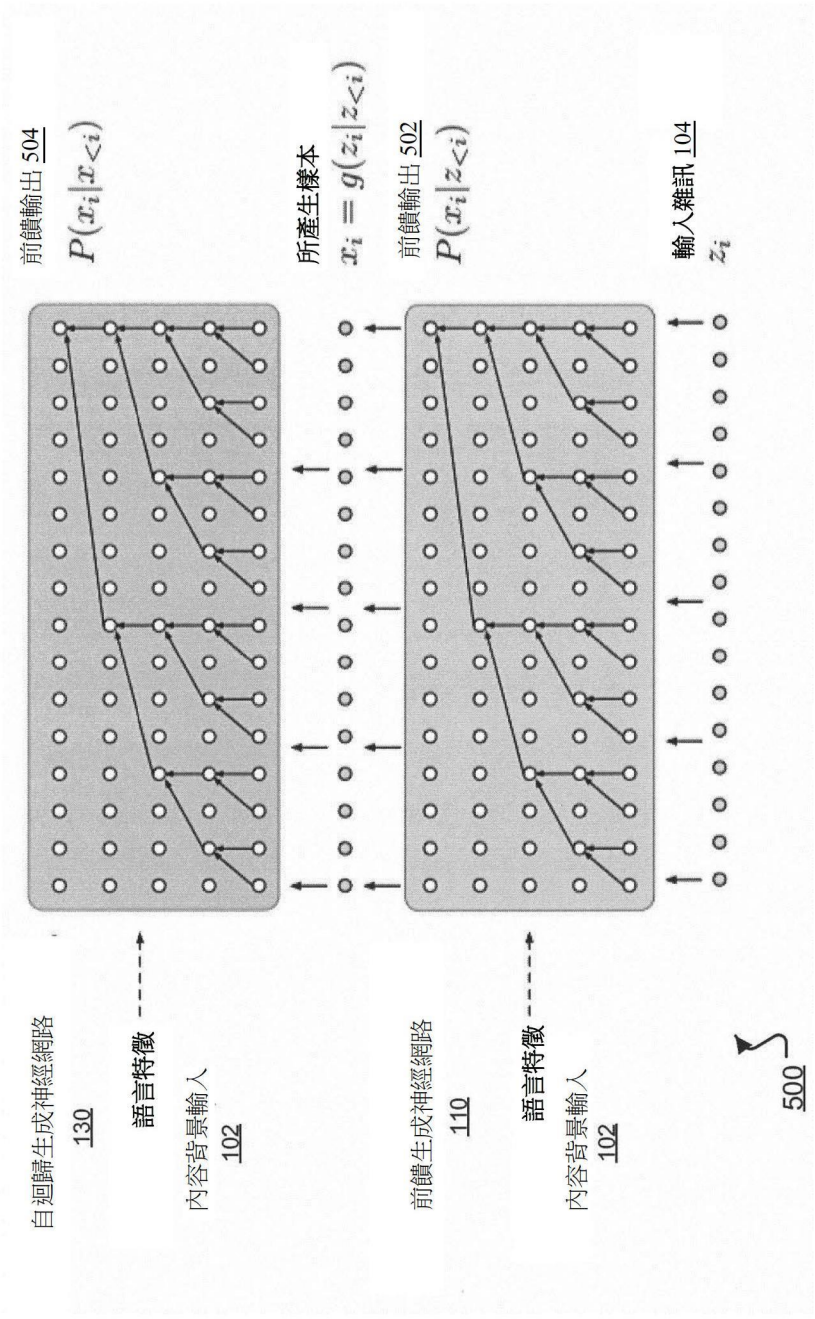
【圖2】



【圖3】


獲得訓練
內容背景輸入
402
使用前饋生成
神經網路處理
404
使用自迴歸生成
神經網路處理
406
判定散度
損耗之梯度
408
判定量值
損耗之梯度
410
判定知覺
損耗之梯度
412
判定對比
損耗之梯度
414
判定參數更新
416
400

【圖4】


自迴歸生成神經網路
130
語言特徵
內容背景輸入
102
前饋輸出 504
$P(x_i|x_{<i})$
所產生樣本
$x_i = g(z_i|z_{<i})$
前饋生成神經網路
110
前饋輸出 502
$P(x_i|z_{<i})$
語言特徵
內容背景輸入
102
輸入雜訊 104
z_i
500

【圖5】