



(19) 中華民國智慧財產局

(12) 發明說明書公開本

(11) 公開編號：TW 201709137 A

(43) 公開日：中華民國 106 (2017) 年 03 月 01 日

(21) 申請案號：104126843

(22) 申請日：中華民國 104 (2015) 年 08 月 18 日

(51) Int. Cl. :

G06Q50/22 (2012.01)

G06F17/16 (2006.01)

(71) 申請人：國立交通大學 (中華民國) NATIONAL CHIAO TUNG UNIVERSITY (TW)

新竹市東區大學路 1001 號

(72) 發明人：林進燈 LIN, CHIN TENG (TW) ; 謝宗佑 HSIEH, TSUNG YU (TW) ; 林洋印 LIN, YANG YIN (TW)

(74) 代理人：林火泉

申請實體審查：有 申請專利範圍項數：8 項 圖式數：2 共 12 頁

(54) 名稱

利用子空間產生取樣之方法

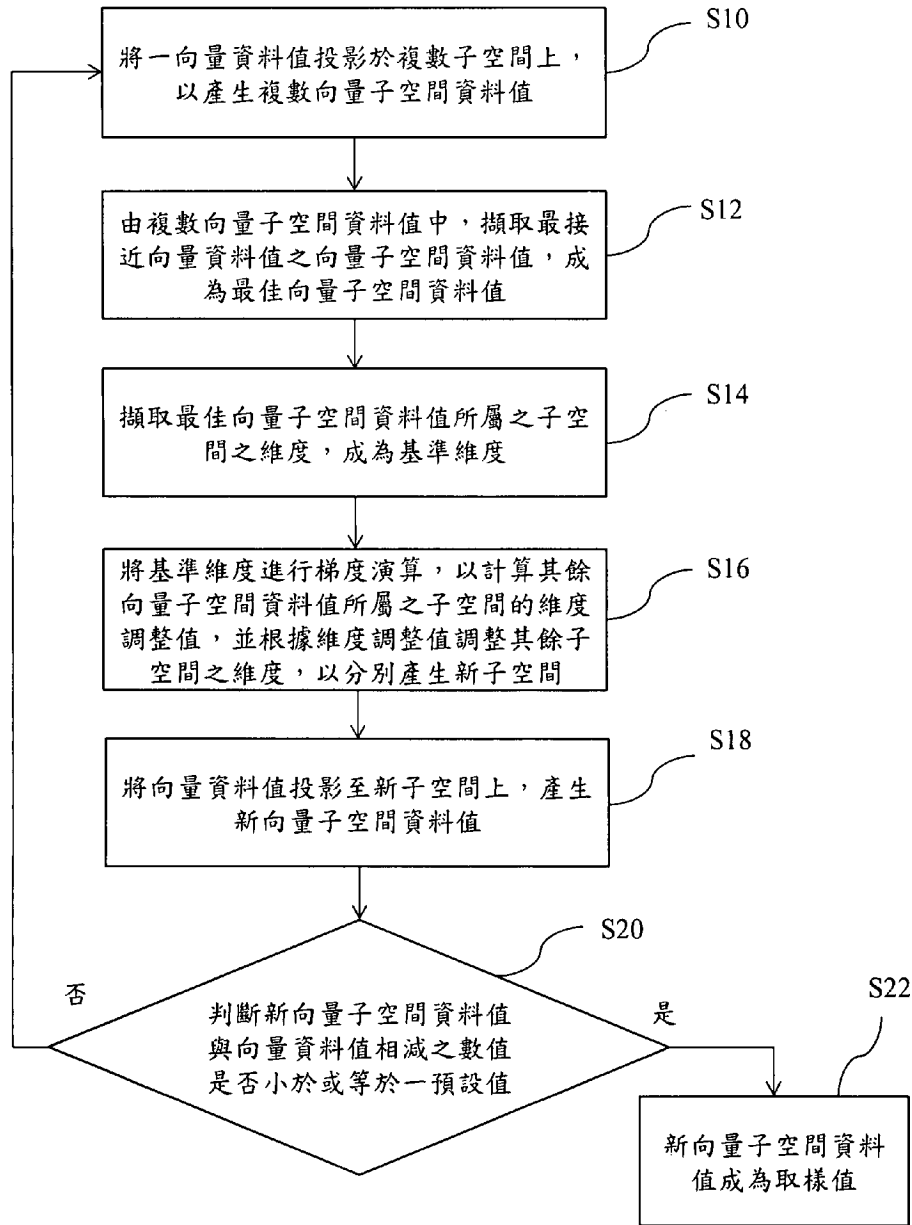
A METHOD OF GENERATED SAMPLING FORM UTILIZING SUBSPACE

(57) 摘要

本發明係一種利用子空間產生取樣之方法，步驟包括將向量資料值投影至複數子空間，產生複數向量子空間資料值，並由其中擷取最接近向量資料值的向量子空間資料值為最佳向量子空間資料值，取其所屬之子空間的維度為基準維度，以利用梯度演算調整其餘子空間之維度，產生新子空間後再投影出新向量子空間資料值，再判斷新向量子空間資料值與向量資料值的差異值是否小於或等於一預設值，若是，新向量子空間資料值則成為取樣值，若否，則回覆至第一步驟。故，本發明可產生與原始資料相似的取樣資料。

The present invention is a method of generated sampling form utilizing subspace, the steps comprising, projecting the vector data on plurality of subspaces, and generating plurality of vector subspace data. Capturing the vector subspace data, that most close the vector data to become the optimum vector subspace data, and capture the dimension of subspace of optimum vector subspace data, and capture optimum vector subspace data into gradient algorithm to adjust other dimension of subspace, and producing a new subspace, then projecting the vector data on new subspace, let the new subspace producing new vector subspace data, then subtract new vector subspace data from vector data get difference value, and to decide the difference value is less or equal to a preset value, if yes, the new vector subspace data will become sampled value, if not, it will return to the first step. The present invention can produce data near the sampling data.

指定代表圖：



第一圖

201709137

【發明摘要】

【中文發明名稱】利用子空間產生取樣之方法

【英文發明名稱】A method of generated sampling form utilizing subspace

【中文】

本發明係一種利用子空間產生取樣之方法，步驟包括將向量資料值投影至複數子空間，產生複數向量子空間資料值，並由其中擷取最接近向量資料值的向量子空間資料值為最佳向量子空間資料值，取其所屬之子空間的維度為基準維度，以利用梯度演算調整其餘子空間之維度，產生新子空間後再投影出新向量子空間資料值，再判斷新向量子空間資料值與向量資料值的差異值是否小於或等於一預設值，若是，新向量子空間資料值則成為取樣值，若否，則回覆至第一步驟。故，本發明可產生與原始資料相似的取樣資料。

【英文】

The present invention is a method of generated sampling form utilizing subspace, the steps comprising, projecting the vector data on plurality of subspaces, and generating plurality of vector subspace data. Capturing the vector subspace data, that most close the vector data to become the optimum vector subspace data, and capture the dimension of subspace of optimum vector subspace data, and capture optimum vector subspace data into gradient algorithm to adjust other dimension of subspace, and producing a new subspace, then projecting the vector data on new subspace, let the new subspace producing new vector subspace data, then subtract new vector subspace data from vector data get difference value, and to decide the difference value is less or equal to a preset value, if yes, the new vector subspace data will become sampled value, if not, it will return to the first step. The present invention can produce data near the sampling data.

【指定代表圖】：第（一）圖。

【代表圖之符號簡單說明】

無

【發明說明書】

【中文發明名稱】 利用子空間產生取樣之方法

【英文發明名稱】 A method of generated sampling form utilizing subspace

【技術領域】

【0001】 本發明係為有關一種產生取樣資料之方法，特別是指一種利用智慧演算法，產生更接近原始資料值之利用子空間產生取樣之方法。

【先前技術】

【0002】 在現今資料科學研究中，不同族群資料間比例失衡所造成的問題已經引起廣大的研究注目，以病患與健康人類來說，通常病患的比例會較健康人類少，若需針對這兩個不同族群做研究時，其資料數應相同，才能較精確地研究出正確的資料，但由於病患之族群人數較少，導致病患之族群的資料明顯不足，此時資料研究的過程就會產生比例失衡的問題，導致研究正確率降低。

【0003】 因此為了解決兩族群之間比例失衡的問題，此時則需自行製造病患之族群的取樣資料，以增加病患族群的取樣資料數目，藉此平衡健康人類以及病患兩個不同族群的數量，但製造取樣的過程中若在非理想的取樣方法下會產生取樣資料失真，導致取樣資料與原始作為基礎的資料相差太多，產生出與原本資料差異太多的取樣資料，導致研究結果正確率降低。再者，若取樣資料與原始之資料差太多，一般製作取樣的方法並無法對取樣資料產生的參數進行調整，導致重複產生錯誤的取樣資料，不但效率低，且錯誤率也相對提高。

【0004】 有鑑於此，本發明遂針對上述習知技術之缺失，提出一種取樣產生方法，以有效克服上述之該等問題。

【發明內容】

【0005】 本發明之主要目的在提供一種利用子空間產生取樣之方法，其係可將原始資料投影至子空間，以產生與原始資料相似的取樣資料，且產生的取

樣資料不但可保有原始資料的特性，並能補足資料庫中資料的不足，以解決不同族群的資料間比例失衡的問題。

【0006】 本發明之另一目的在提供一種利用子空間產生取樣之方法，其係可結合梯度演算法調整子空間之維度，藉由不斷調整子空間之維度，產生更接近原始資料的取樣資料。

【0007】 為達上述之目的，本發明提供一種利用子空間產生取樣之方法，包括下列步驟，利用一計算機將一向量資料值投影至複數子空間上，以產生複數向量空間資料值；由複數向量子空間資料值中擷取最接近向量資料值的向量子空間資料值，為最佳向量子空間資料值；接著擷取最佳向量子空間資料值所屬子空間的維度為基準維度；將基準維度進行梯度演算，以計算其餘向量子空間值所屬子空間的維度調整值，並根據維度調整值調整其餘子空間的維度，以分別產生一個新子空間；將向量資料值投影至每一新子空間上，產生複數新向量子空間資料值；最後判斷每一新向量子空間資料值與向量資料值相減之數值是否小於或等於一預設值，若是，新向量子空間資料值則成為取樣值，但若否，則回覆至將一向量資料值投影於複數子空間之步驟，並重複進行上述步驟。

【0008】 底下藉由具體實施例詳加說明，當更容易瞭解本發明之目的、技術內容、特點及其所達成之功效。

【圖式簡單說明】

【0009】

第一圖係為本發明之方法流程圖。

第二圖係為本發明之向量資料值與向量子空間資料值分布示意圖。

【實施方式】

【0010】 本發明可應用於計算機中，如電腦等可供計算之工具，接著介紹本發明之方法應用於計算機時的步驟，請配合參照第一圖，其係為本發明之方

法流程圖，如圖所示，首先進入步驟S10，先定義所需子空間的數目、維度以及空間參數後，將一向量資料值投影至複數子空間上，以產生複數向量空間資料值。接著進入步驟S12，由複數向量子空間資料值中擷取最接近向量資料值的向量子空間資料值，為最佳向量子空間資料值，其中上述取得最佳向量子空間資料值的方法係利用一最佳向量子空間資料值方程式(1)取得，其中最佳向量子空間資料值方程式(1)如下所示：

$$c = \operatorname{argmin} \left\| x - \hat{x}^i \right\|^2 \dots\dots\dots(1)$$

其中c係為最佳向量子空間資料值，x係為向量資料值， \hat{x}^i 係為向量子空間資料值。因此向量資料值係分別與每一向量子空間資料值相減，產生出複數向量距離值後，擷取最短的向量距離值，成為最佳向量子空間資料值，由於距離越短表示向量資料值與向量子空間資料值的值會最接近，因此本實施例係由複數向量距離值中擷取最小的向量距離值，為最佳向量子空間資料值。

【0011】 接下來進入步驟S14，擷取最佳向量子空間資料值所屬之子空間之維度，成為基準維度。如步驟S16，將基準維度進行梯度演算，以計算其餘未被選擇為最佳向量子空間資料值的複數向量子空間值所屬子空間的維度調整值，並根據維度調整值調整其餘每一子空間的維度，以分別產生一個新子空間，其中梯度演算所使用之方程式(2)如下所示：

$$b_h^{(i)}(t+1) = \left[I + \lambda(t) h_c^{(i)} \left(t \frac{x(t)x(t)^T}{\|\hat{x}^{(i)}(t)\| \|x(t)\|} \right) \right] b_h^{(i)}(t) \dots\dots\dots(2)$$

其中 $b_h^{(i)}$ 係為子空間，t係為目前所調整之子空間，I係為一預設矩陣， $\lambda(t)$ 係為一常數， $h_c^{(i)}$ 係為維度調整值，x係為向量資料值， \hat{x} 係為向量子空間資料值。因此即可利用梯度演算所使用之方程式(2)可逐一調整其餘每一子空間的維度，將剩餘的子空間調整成與最佳向量子空間資料值的子空間相似維度的子空間，以於投影時產生接近向量資料值的向量子空間資料值。

【0012】 接下來進入步驟S18，將向量資料值投影至調整過後的新子空間上，以產生新向量子空間資料值，由於新子空間係依據最佳向量子空間資料值的子空間維度去調整的，因此新子空間所投影出來的新向量子空間資料值係為接近向量資料值的值。最後進入步驟S20，利用一方程式(3)將新向量子空間值與向量資料值相減，並將新向量子空間值與向量資料值相減之數值是否小於或等於一預設值，以再次確認新向量子空間資料值的值是否夠接近向量資料值，其中上述之方程式(3)如下所示：

$$\sum_n |x_n - \widehat{x}_n^c|^2 \leq \varepsilon \dots \dots \dots (3)$$

其中 x_n 係為向量資料值，其中 \widehat{x}_n^c 係為新向量子空間資料值，其中 ε 係為預設值。本實施例舉例預設值為0.1，因此若差異值小於或等於0.1則表示計算出來的新向量子空間資料值係接近向量資料值，可被接受，故可進入步驟S22，將新向量子空間資料值成為取樣值，以將取樣值儲存至一取樣資料庫中，以增加資料庫內部的取樣資料數目，以作為研究數據；但若為否，差異值不是小於或等於0.1，則回覆步驟S10，以重新將向量資料值投影於複數子空間，以重新產生並計算新向量子空間資料值。

【0013】 接下來請參照第二圖，其係為向量資料值與向量子空間資料值分布示意圖，其中圓形的點係為向量資料值，X形的點係為向量子空間資料值，由第二圖可知，本發明產生出與原始的向量資料值相近的向量子空間資料值，以補足原本資料庫缺少的數目平衡不同族群之間的樣本數，提供較佳的實驗數據。

【0014】 綜上所述，本發明可將原始資料投影至子空間中，以產生與原始資料相似的取樣資料，產生出來的取樣資料可保有原始資料的特性，並且能補足原本資料庫中資料不足，以解決不同族群的資料間比例失衡的問題，且本發明結合梯度演算法來調整子空間之維度，藉由不斷的演算來調整子空間之維度，以產生更接近原始資料的取樣資料。

【0015】 唯以上所述者，僅為本發明之較佳實施例而已，並非用來限定本發明實施之範圍。故即凡依本發明申請範圍所述之特徵及精神所為之均等變化或修飾，均應包括於本發明之申請專利範圍內。

【符號說明】

【0016】 無

【發明申請專利範圍】

【第1項】 一種利用子空間產生取樣之方法，包括下列步驟：

利用一計算機將一向量資料值投影於複數子空間上，以產生複數向量子空間資料值；

由該等向量子空間資料值中，擷取最接近該向量資料值之該向量子空間資料值，成為最佳向量子空間資料值；

擷取該最佳向量子空間資料值所屬之該子空間之維度，成為基準維度；

將該基準維度進行一梯度演算，以計算其餘該向量子空間資料值所屬之該子空間的維度調整值，並根據該維度調整值調整其餘該子空間之維度，以分別產生一新子空間；

將該向量資料值投影至每一該新子空間上，產生複數新向量子空間資料值；以及

判斷每一該新向量子空間資料值與該向量資料值相減之數值是否小於或等於一預設值；

若是，該新向量子空間資料值則成為取樣值；及

若否，則回覆至將該向量資料值投影於該等子空間之步驟，並重複進行上述步驟。

【第2項】 如請求項1所述之利用子空間產生取樣之方法，其中在擷取最接近該向量資料值之該向量子空間資料值之步驟中，係將該向量資料值分別與每一該向量子空間資料值相減，產生複數向量距離值，並擷取最小的向量距離值，為該最佳向量子空間資料值。

【第3項】 如請求項2所述之利用子空間產生取樣之方法，其中在擷取最接近該向量資料值之該向量子空間資料值之步驟中，係透過一最佳向量子

空間資料值方程式取得該最佳向量子空間資料值，該最佳向量子空間資料值方程式如下所示：

$$c = \operatorname{argmin} \left\| x - \hat{x}^i \right\|^2$$

其中該 c 係為該最佳向量子空間資料值，該 x 係為該向量資料值，該 \hat{x}^i 係為該向量子空間資料值。

【第4項】如請求項1所述之利用子空間產生取樣之方法，其中該梯度演算使用之方程式為：

$$b_h^{(i)}(t+1) = \left[I + \lambda(t) h_c^{(i)} \left(t \frac{x(t)x(t)^T}{\|\hat{x}^{(i)}(t)\| \|x(t)\|} \right) \right] b_h^{(i)}(t)$$

其中該 $b_h^{(i)}$ 係為該子空間，該 t 係為目前所調整之子空間，該 I 係為一預設矩陣，該 $\lambda(t)$ 係為一常數，該 $h_c^{(i)}$ 係為維度調整值，該 x 係為該向量資料值，該 \hat{x} 係為該向量子空間資料值。

【第5項】如請求項1所述之利用子空間產生取樣之方法，其中在將該向量資料值投影於該等子空間上之步驟之前，可定義該子空間數目、維度以及空間參數。

【第6項】如請求項1所述之利用子空間產生取樣之方法，其中在成為該取樣值之步驟之後，更包括將該取樣值儲存至一取樣資料庫中，以增加該資料庫之取樣資料。

【第7項】如請求項1所述之利用子空間產生取樣之方法，其中在判斷每一該新向量子空間資料值與該向量資料值相減之數值是否小於或等於該預設值之步驟更可利用下列方程式進行判斷：

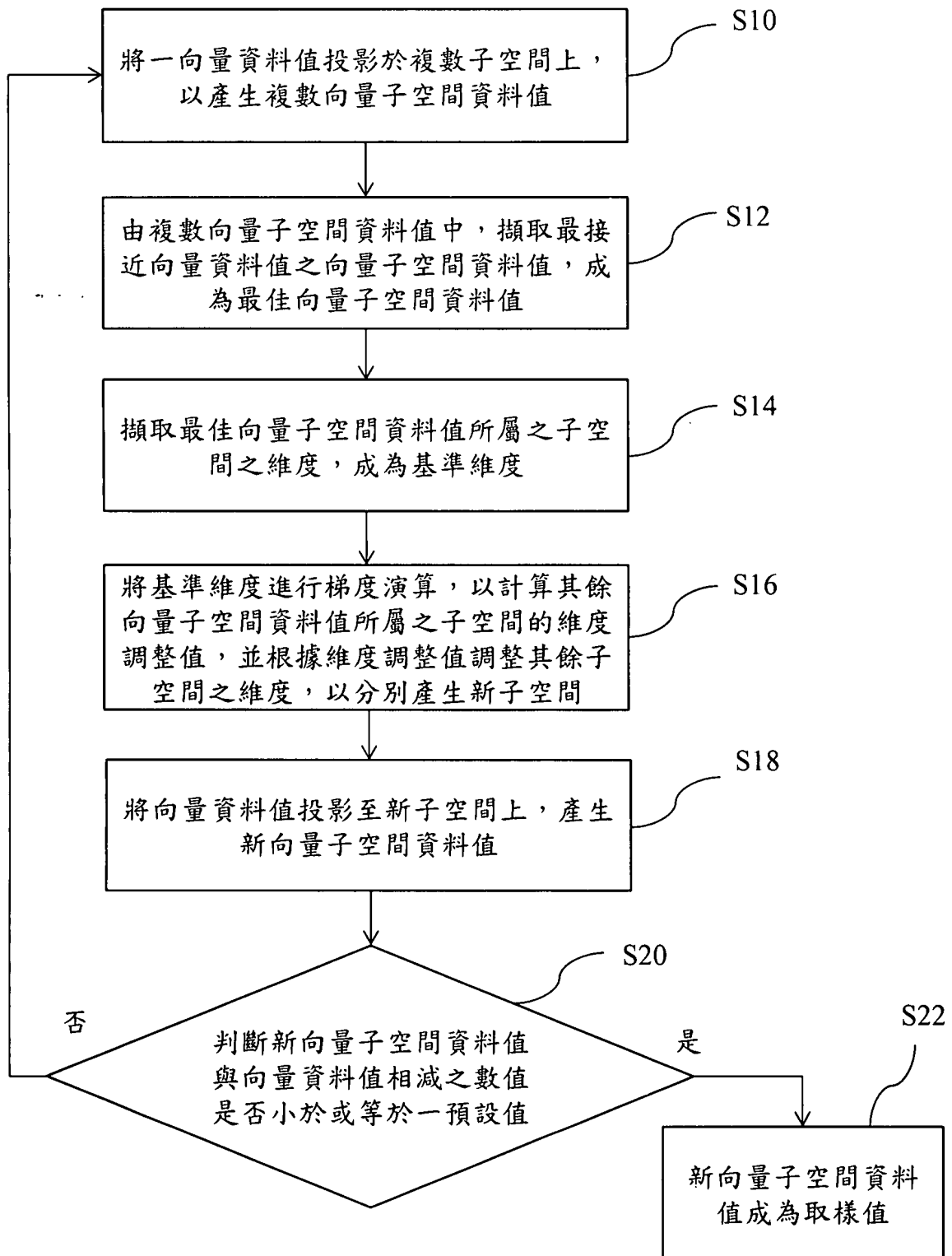
$$\sum_n \left\| x_n - \hat{x}_n^c \right\|^2 \leq \varepsilon$$

其中該 x_n 係為該向量資料值，其中該 \hat{x}_n^c 係為該新向量子空間資料

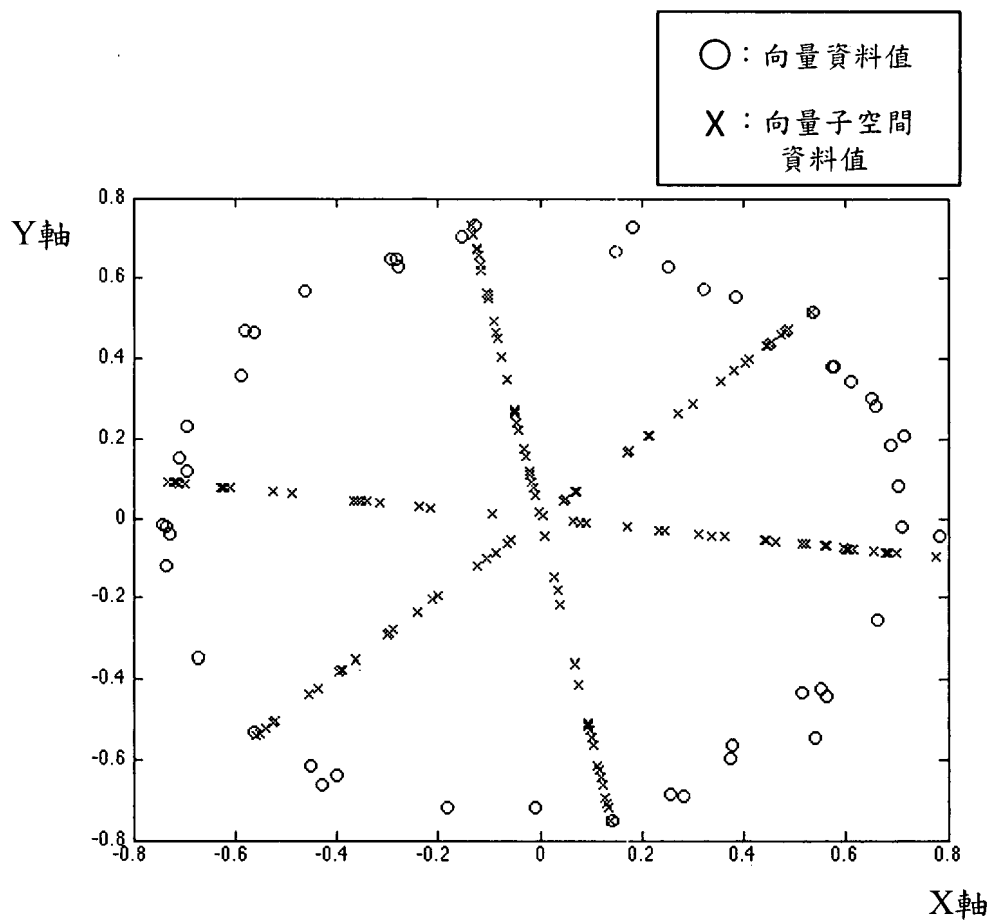
值，其中該 ϵ 係為該預設值。

【第8項】 如請求項7所述之利用子空間產生取樣之方法，其中該預設值係為0.1。

【發明圖式】



第一圖



第二圖