



(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 201709086 A

(43) 公開日：中華民國 106 (2017) 年 03 月 01 日

(21) 申請案號：104126844

(22) 申請日：中華民國 104 (2015) 年 08 月 18 日

(51) Int. Cl. :

*G06F17/18 (2006.01)**G06F17/10 (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)；劉宇庭 LIU, YU TING (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

利用高維度子空間產生取樣之方法

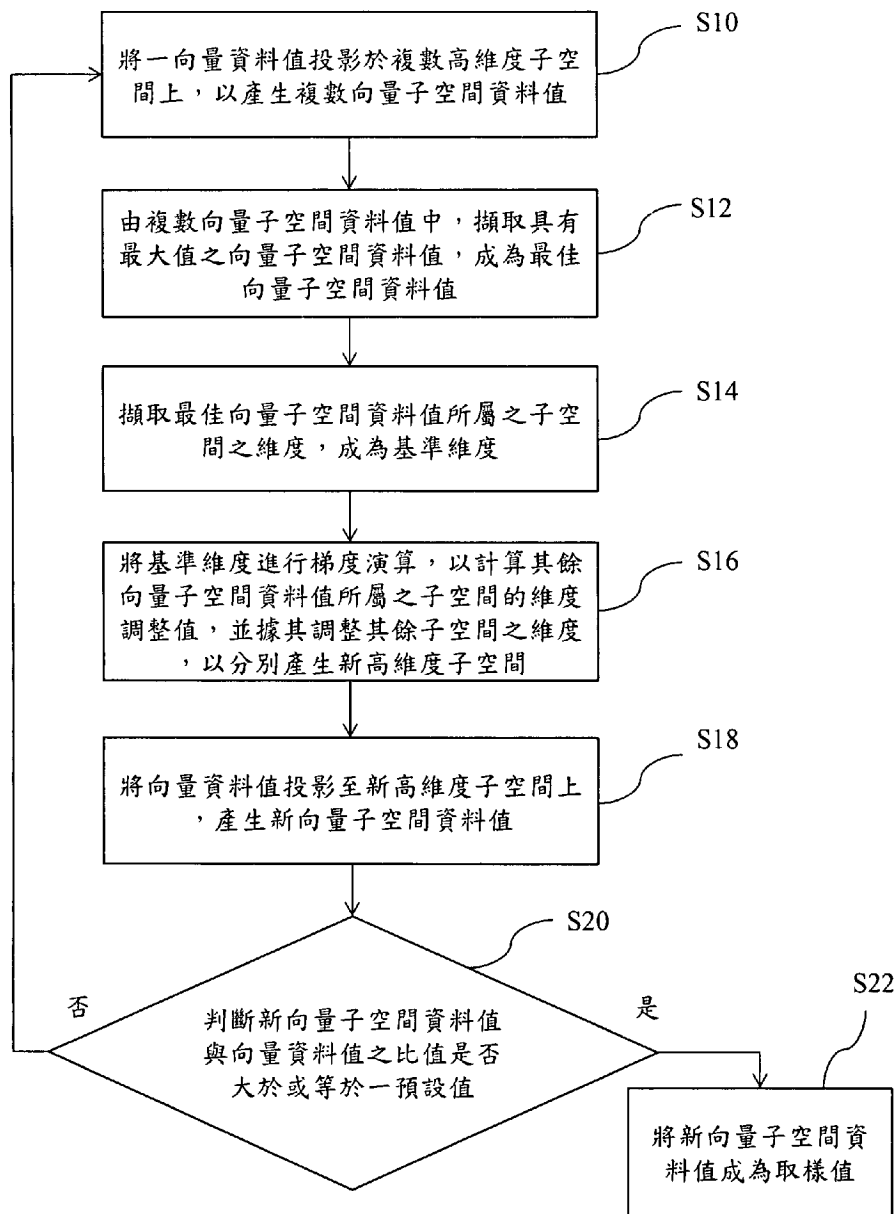
METHOD OF GENERATED SAMPLING FORM UTILIZING KERNEL SUBSPACE

(57) 摘要

本發明係為一種利用高維度子空間產生取樣之方法，包括將一向量資料值投影至複數高維度子空間上，以產生複數向量子空間資料值，並取其中具有最大值之向量子空間資料值為最佳向量子空間資料值，並取其所屬子空間的維度，設為基準維度並利用梯度演算，調整其餘子空間，以產生新高維度子空間，再投影出新向量子空間資料值，判斷新向量子空間資料值與向量資料值之比值是大於或等於一預設值，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 kernel subspace, the steps comprising, projecting the vector data on plurality of kernel subspaces, and generating plurality of vector subspace data. Capturing the max data of vector subspace data to become the optimum vector subspace data, and capture the dimension of subspace of optimum vector subspace data, and capturing optimum vector subspace data into gradient algorithm to adjust other dimension of kernel subspace, and producing a new kernel subspace, then projecting the vector data on new kernel subspace, let the new kernel subspace producing a new vector subspace data, and then to decide the ratio of new vector subspace data to vector data is equal or greater than 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

201709086

## 【發明摘要】

G06F 17/18 (2006.01)

【中文發明名稱】利用高維度子空間產生取樣之方法

G06F 17/10 (2006.01)

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

## 【中文】

本發明係為一種利用高維度子空間產生取樣之方法，包括將一向量資料值投影至複數高維度子空間上，以產生複數向量子空間資料值，並取其中具有最大值之向量子空間資料值為最佳向量子空間資料值，並取其所屬子空間的維度，設為基準維度並利用梯度演算，調整其餘子空間，以產生新高維度子空間，再投影出新向量子空間資料值，判斷新向量子空間資料值與向量資料值之比值是大於或等於一預設值，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 kernel subspace, the steps comprising, projecting the vector data on plurality of kernel subspaces, and generating plurality of vector subspace data. Capturing the max data of vector subspace data to become the optimum vector subspace data, and capture the dimension of subspace of optimum vector subspace data, and capturing optimum vector subspace data into gradient algorithm to adjust other dimension of kernel subspace, and producing a new kernel subspace, then projecting the vector data on new kernel subspace, let the new kernel subspace producing a new vector subspace data, and then to decide the ratio of new vector subspace data to vector data is equal or greater than 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】 利用高維度子空間產生取樣之方法

【英文發明名稱】 method of generated sampling form utilizing kernel 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，以產生更接近原始資料值之利用高維度子空間產生取樣之方法。

【先前技術】

【0002】 目前的資料科學研究中，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，以醫學的領域，並以病患與健康人類為例，通常病患的比例會較健康人類少，因此病患可取樣的數據相對於健康人類來說，可取樣的數據會比健康人類少很多。且在針對這兩個不同族群做研究時，其資料樣本數應相同，才能較精確地比較兩者的不同，以研究出正確的資料，但若在研究時，病患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，導致研究的準確度降低。

【0003】 因此為了提高研究的準確度以及解決兩族群之間比例失衡的問題，必須自行產生病患族群的取樣資料，以增加病患族群的取樣資料數目，來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，若在非理想狀態可能導致產生的取樣資料失真，使取樣資料的數值與原始作為基礎資料數值相差太多，以至於資料不正確，導致研究結果準確度降低。再者，先前技術在取樣資料與原始之資料差太多時，也無法針對取樣資料進行調整，導致重複產生錯誤的取樣資料，使研究產生錯誤。

【0004】 有鑑於此，本發明遂針對上述習知技術之缺失，提出一種利用高維度子空間產生取樣之方法，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高維度子空間產生取樣之方法，其可將原始資料投影至高維度子空間，以產生與原始資料相似的取樣資料，且高維度子空間下所產生的取樣資料可與原始資料的特性幾乎完全相同，故能做為平衡數據之資料，以補足資料庫中資料的不足，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高維度子空間產生取樣之方法，其係可結合梯度演算法調整高維度子空間之維度，藉由不斷調整高維度子空間之維度，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高維度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高維度子空間上，以產生複數向量子空間資料值；由複數向量子空間資料值中，擷取具有最大值之向量子空間資料值，成為最佳向量子空間資料值；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；將基準維度進入一梯度演算，以計算其餘的向量子空間資料值所屬之子空間的維度調整值，並根據維度調整值調整其餘子空間之維度，以分別產生一新的高維度子空間；將向量資料值投影至每一新高維度子空間上，產生複數新向量子空間資料值；判斷每一新向量子空間資料值與向量資料值之比值是否大於或等於一預設值，若是，新向量子空間資料值則成為取樣值，若否，則回覆至將一向量資料值投影於複數高維度子空間之步驟，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，請參照第一圖，以說明本發明應用於計算機時的演算方法及步驟，如圖所示，首先進入步驟S10，操作計算機以調整所需高維度子空間的數目、維度以及空間參數後，將一向量資料值投影於複數高維度子空間上，以產生複數向量子空間資料值。接著進入步驟S12，由複數向量子空間資料值中，擷取具有最大值的向量子空間資料值，成為最佳向量子空間資料值，其中擷取具有最大值之向量子空間資料值的步驟演算係透過一最佳向量子空間資料值方程式(1)，取得最大值的向量子空間資料值成為最佳向量子空間資料值，其中最佳向量子空間資料值方程式(1)如下所示：

$$C = \operatorname{argmax}_{i \in \text{module}} \left\| \widehat{\varphi}_i(x) \right\|^2 \dots\dots\dots(1)$$

其中C係為最佳向量子空間資料值， $\widehat{\varphi}_i(x)$ 係為向量子空間資料值。

**【0011】** 經由上述方程式(1)取得最佳向量子空間資料值後，接著進入步驟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。接著進入步驟S16，將基準維度進行梯度演算，以逐一計算其餘向量子空間資料值所屬之子空間的維度調整值，使每一向量子空間的維度可根據維度調整值進行調整，分別產生出一新高維度子空間，以利投影出較佳的向量子空間資料值，其中梯度演算使用之方程式(2)如下所示：

$$\alpha_{i,h,n}^{(L+1)} = \alpha_{i,h,n}^{(L)} + \sigma \frac{\partial D(x_j)}{\partial \alpha_{i,h,n}} \dots\dots\dots(2)$$

其中 $\alpha_{i,h,n}^{(L)}$ 係為子空間，L係為目前所調整之子空間， $\sigma$ 係為常數參數， $D(x_j)$ 係為維度調整值。

**【0012】** 接著進入步驟S18，將向量資料值投影至每一新高維度子空間

上，產生複數新向量子空間資料值。進入步驟S20，計算每一新向量子空間資料值與向量資料值之比值，其中計算新向量子空間資料值與向量資料值之比值係利用方程式(3)計算，方程式(3)如下所示：

$$\frac{\sum_n \frac{\|\hat{\varphi}^c(x_n)\|^2}{\|\varphi(x_n)\|^2}}{N} \geq \varepsilon \dots \dots \dots (3)$$

其中 $\varphi(x_n)$ 係為向量資料值， $\hat{\varphi}^c(x_n)$ 係為新向量子空間資料值， $\varepsilon$ 係為預設值， $N$ 係為新向量子空間資料值的總數。計算出比值後並判斷是否大於或等於一預設值，本實施例舉例預設值係為0.98，若是，計算出來的比值大於或等於0.98，則進入步驟S22，將新向量子空間資料值成為取樣值，並將取樣值儲存至一資料庫中，以增加資料庫的取樣資料；但若否，則回覆至步驟S10，重複將向量資料值投影於複數高維度子空間之，以計算出更精確的向量子空間資料值。

**【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，利用本發明之方法可產生出與原始的向量資料值相當相近的向量子空間資料值，產生出的向量子空間資料值加入資料庫中，以補足原本資料庫缺少的數目，來平衡不同族群之間的樣本數，藉此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高維度子空間，以產生與原始資料相似的取樣資料，且高維度子空間下所產生的取樣資料可與原始資料的特性幾乎完全相同，故做為補足資料庫中不足的資料，以解決不同族群的資料間比例失衡的問題。除此之外，本發明更結合梯度演算法，來調整高維度子空間之維度，藉由不斷調整高維度子空間之維度，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】 一種利用高維度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高維度子空間上，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，擷取具有最大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，以分別產生一新高維度子空間；
- 將該向量資料值投影至每一該新高維度子空間上，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之比值是否大於或等於一預設值：
- 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高維度子空間之步驟，以重覆進行上述之步驟。
- 【第2項】 如請求項1所述之利用高維度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，係由該等向量子空間資料值中擷取具最大值的該向量子空間資料，為該最佳向量子空間資料值。
- 【第3項】 如請求項2所述之利用高維度子空間產生取樣之方法，其中在擷取具有最大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：

$$C = \operatorname{argmax}_{i \in \text{subspace}} \|\hat{\varphi}_i(x)\|^2$$

其中該C係為該最佳向量子空間資料值，該 $\hat{\varphi}_i(x)$ 係為該向量子空間資料值。

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

$$\alpha_{i,h,n}^{(L+1)} = \alpha_{i,h,n}^{(L)} + \sigma \frac{\partial D(x_j)}{\partial \alpha_{i,h,n}}$$

其中 $\alpha_{i,h,n}^{(L)}$ 係為該子空間，該L係為目前所調整之子空間，該 $\sigma$ 係為常數參數，該 $D(x_j)$ 係為該維度調整值。

【第5項】如請求項1所述之利用高維度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高維度子空間上之步驟之前，可定義高維度子空間數目、維度以及空間參數。

【第6項】如請求項1所述之利用高維度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。

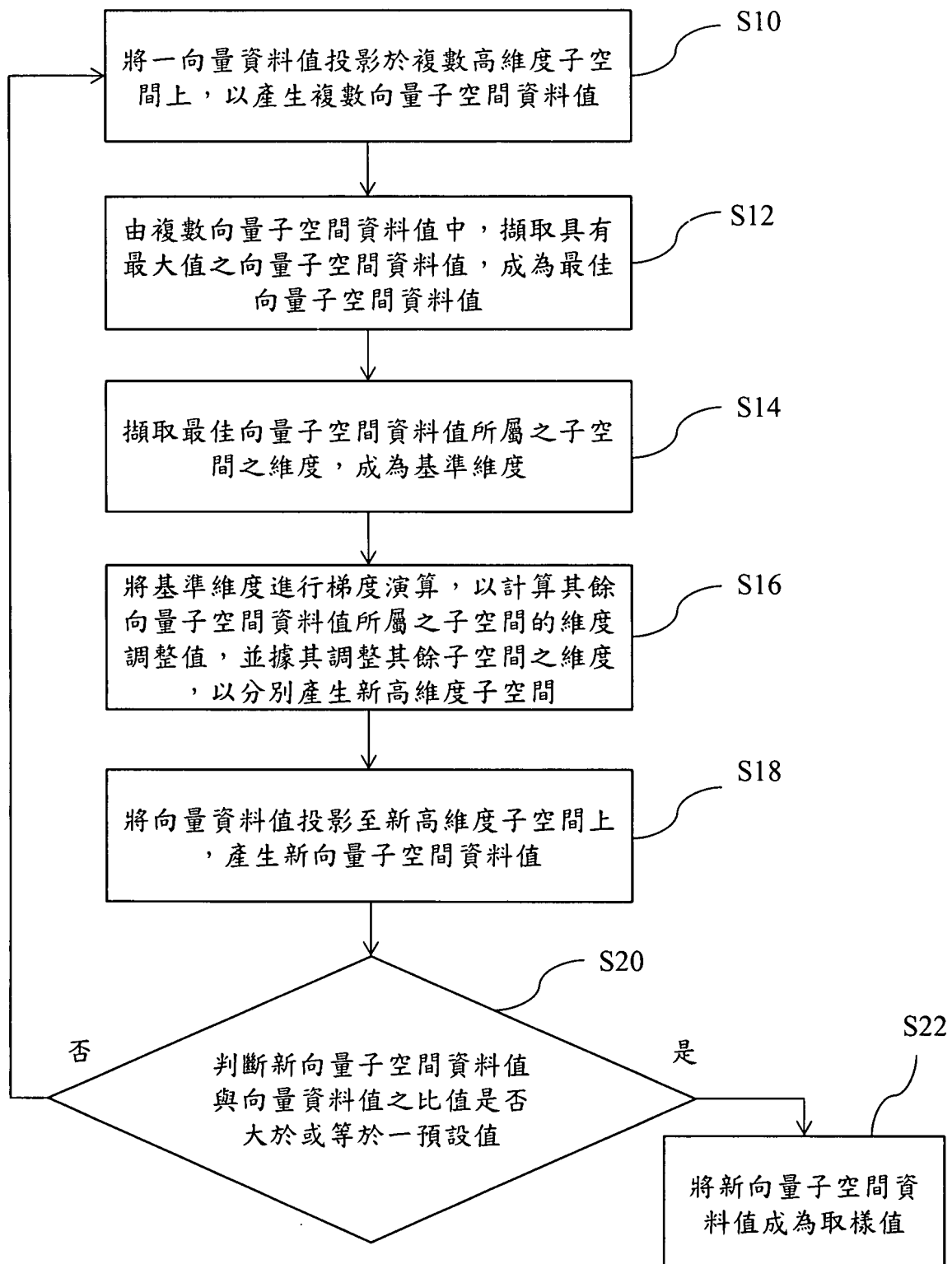
【第7項】如請求項1所述之利用高維度子空間產生取樣之方法，其中在計算該新向量子空間資料值與該向量資料值之比值是否大於或等於該預設值之步驟中，係利用下列方程式進行判斷：

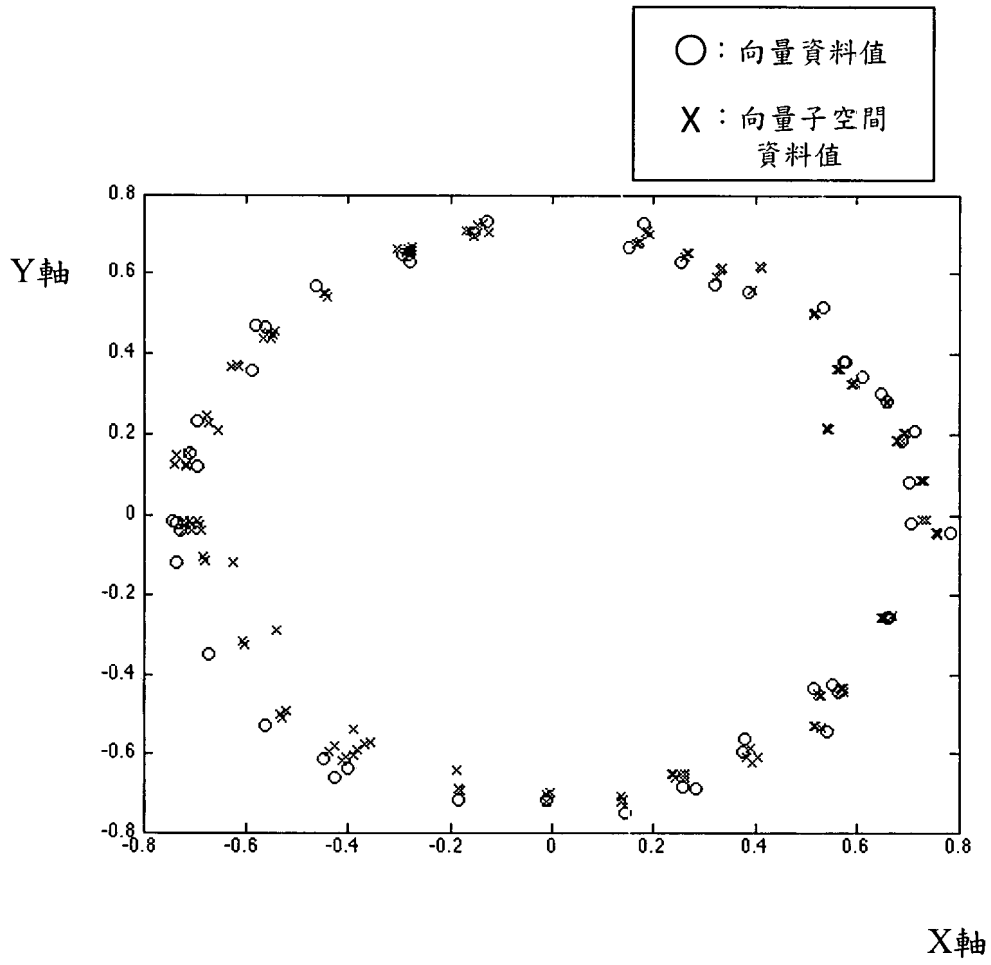
$$\frac{\sum_n \frac{\|\hat{\varphi}^c(x_n)\|^2}{\|\varphi(x_n)\|^2}}{N} \geq \varepsilon$$

其中該 $\varphi(x_n)$ 係為該向量資料值，其中該 $\hat{\varphi}^c(x_n)$ 係為新向量子空間資料值，其中該 $\varepsilon$ 係為該預設值，其中該N係為該新向量子空間資料值的總數。

【第8項】 如請求項7所述之利用高維度子空間產生取樣之方法，其中該預設值係為0.98。

## 【發明圖式】





第二圖