I718643

(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 I718643 B**

(45)公告日： 中華民國 110 (2021) 年 02 月 11 日

(21)申請案號：108130766　(22)申請日： 中華民國 108 (2019) 年 08 月 28 日

(51)Int. Cl.：　*G06Q10/06　(2012.01)*　*G06Q20/40　(2012.01)*

(30)優先權：2019/01/17　中國大陸　201910045152.6

(71)申請人：開曼群島商創新先進技術有限公司 (開曼群島) ADVANCED NEW TECHNOLOGIES CO., LTD.　(KY)

開曼群島

(72)發明人：苗加成 (CN) ；章鵬 (CN) ；楊程遠 (CN) ；向彪 (CN) ；嚴歡 (CN)

(74)代理人：林志剛

(56)參考文獻：

TW	I560577	CN	103812872A
CN	103927398B	CN	107332931A
CN	107391548A	CN	107870934A
US	8719190B2		

審查人員：徐瑞甫

申請專利範圍項數：11 項　圖式數：8　共 63 頁

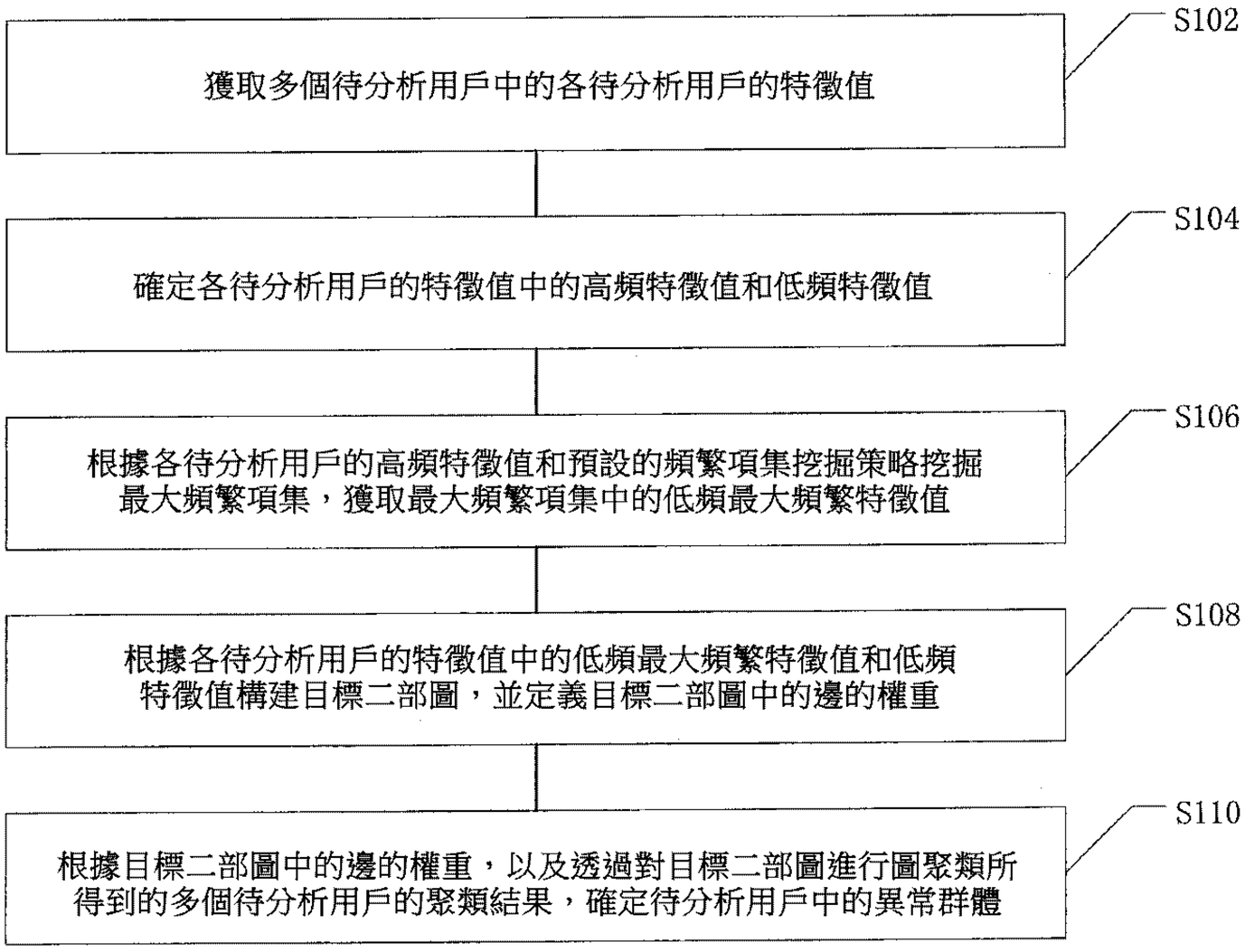
(54)名稱

異常群體識別方法及裝置

(57)摘要

本發明實施例提供了一種異常群體識別方法及裝置。其中方法包括：獲取多個待分析用戶中的各待分析用戶的特徵值；確定各待分析用戶的特徵值中的高頻特徵值和低頻特徵值；根據各待分析用戶的高頻特徵值和預設的頻繁項集挖掘策略挖掘最大頻繁項集，獲取最大頻繁項集中的低頻最大頻繁特徵值；根據各待分析用戶的特徵值中的低頻最大頻繁特徵值和低頻特徵值構建目標二部圖，並定義目標二部圖中的邊的權重；根據目標二部圖中的邊的權重，以及透過對目標二部圖進行圖聚類所得到的多個待分析用戶的聚類結果，確定待分析用戶中的異常群體。本發明實施例提高了異常群體識別的準確率，且步驟簡單，易於執行。

指定代表圖：


獲取多個待分析用戶中的各待分析用戶的特徵值
S102
確定各待分析用戶的特徵值中的高頻特徵值和低頻特徵值
S104
根據各待分析用戶的高頻特徵值和預設的頻繁項集挖掘策略挖掘最大頻繁項集，獲取最大頻繁項集中的低頻最大頻繁特徵值
S106
根據各待分析用戶的特徵值中的低頻最大頻繁特徵值和低頻特徵值構建目標二部圖，並定義目標二部圖中的邊的權重
S108
根據目標二部圖中的邊的權重，以及透過對目標二部圖進行圖聚類所得到的多個待分析用戶的聚類結果，確定待分析用戶中的異常群體
S110

【圖 1】

專利案號：108130766

公告本

I718643

【發明摘要】

【中文發明名稱】

異常群體識別方法及裝置

【中文】

本發明實施例提供了一種異常群體識別方法及裝置。其中方法包括：獲取多個待分析用戶中的各待分析用戶的特徵值；確定各待分析用戶的特徵值中的高頻特徵值和低頻特徵值；根據各待分析用戶的高頻特徵值和預設的頻繁項集挖掘策略挖掘最大頻繁項集，獲取最大頻繁項集中的低頻最大頻繁特徵值；根據各待分析用戶的特徵值中的低頻最大頻繁特徵值和低頻特徵值構建目標二部圖，並定義目標二部圖中的邊的權重；根據目標二部圖中的邊的權重，以及透過對目標二部圖進行圖聚類所得到的多個待分析用戶的聚類結果，確定待分析用戶中的異常群體。本發明實施例提高了異常群體識別的準確率，且步驟簡單，易於執行。

【指定代表圖】第(1)圖。
【代表圖之符號簡單說明】無

【特徵化學式】無

【發明說明書】

【中文發明名稱】

異常群體識別方法及裝置

【技術領域】

本說明書有關電腦技術領域，尤其有關一種異常群體識別方法及裝置。

【先前技術】

目前，在風控領域中的各種場景(如垃圾註冊、行銷作弊、盜卡盜帳號、騙保等)中，團夥作案的趨勢越來越明顯，嚴重的影響了正常的商業秩序，給商家造成了巨大的損失。因此，如何識別團夥(即異常群體)已經成為商家在運營過程中的重要問題之一。

在常用的異常群體的識別方式中，由於標籤樣本的缺失和異常群體作案方式的多變性，導致異常群體識別準確率較低。

【發明內容】

本說明書一個或多個實施例的目的是提供一種異常群體識別方法及裝置，用以解決現有技術中異常群體識別準確率較低的問題。

為解決上述技術問題，本說明書一個或多個實施例是這樣實現的：

一方面，本說明書一個或多個實施例提供一種異常群體識別方法，包括：

獲取多個待分析用戶中的各所述待分析用戶的特徵值；

確定各所述待分析用戶的特徵值中的高頻特徵值和低頻特徵值；

根據各所述待分析用戶的高頻特徵值和預設的頻繁項集挖掘策略挖掘最大頻繁項集，獲取所述最大頻繁項集中的低頻最大頻繁特徵值；

根據各所述待分析用戶的特徵值中的所述低頻最大頻繁特徵值和所述低頻特徵值構建目標二部圖，並定義所述目標二部圖中的邊的權重；

根據所述目標二部圖中的邊的權重，以及透過對所述目標二部圖進行圖聚類所得到的所述多個待分析用戶的聚類結果，確定所述待分析用戶中的異常群體。

可選地，所述獲取多個待分析用戶中的各所述待分析用戶的特徵值包括：

獲取所述多個待分析用戶的原始個人資料；

對所述多個待分析用戶的原始個人資料進行離散化，以得到各所述待分析用戶的特徵值。

可選地，所述確定各所述待分析用戶的特徵值中的高頻特徵值和低頻特徵值包括：

根據各所述待分析用戶的特徵值構建第一二部圖，其中，所述第一二部圖包括與各所述待分析用戶對應的節

點、與各所述特徵值對應的節點、以及各所述待分析用戶對應的節點與其特徵值對應的節點之間的邊；

在所述第一二部圖中獲取各所述特徵值對應的節點的度，並根據各所述特徵值對應的節點的度在所述特徵值中確定高頻特徵值和低頻特徵值；

根據所述高頻特徵值和所述低頻特徵值確定各所述待分析用戶的特徵值中的高頻特徵值和低頻特徵值。

可選地，所述根據各所述待分析用戶的高頻特徵值和預設的頻繁項集挖掘策略挖掘最大頻繁項集，獲取所述最大頻繁項集中的低頻最大頻繁特徵值包括：

根據各所述待分析用戶的高頻特徵值並結合FP-Growth方法，挖掘支援度滿足預設支援度的頻繁多項集，並在所述頻繁多項集中確定最大頻繁項集；

將各所述待分析用戶的特徵值與所述最大頻繁項集中的最大頻繁特徵值進行匹配，以得到各所述待分析用戶的最大頻繁特徵值；

在所述待分析用戶的最大頻繁特徵值中確定低頻最大頻繁特徵值。

可選地，所述在所述待分析用戶的最大頻繁特徵值中確定低頻最大頻繁特徵值包括：

根據各所述待分析用戶的最大頻繁特徵值構建第二二部圖，其中，所述第二二部圖包括與各所述待分析用戶對應的節點、與各所述最大頻繁特徵值對應的節點、以及各所述待分析用戶對應的節點與其最大頻繁特徵值對應的節

點之間的邊；

在所述第二二部圖中獲取各所述最大頻繁特徵值對應的節點的度，並根據各所述最大頻繁特徵值對應的節點的度在所述最大頻繁特徵值中確定低頻最大頻繁特徵值。

可選地，所述根據所述目標二部圖中的邊的權重，以及透過對所述目標二部圖進行圖聚類所得到的所述多個待分析用戶的聚類結果，確定所述待分析用戶中的異常群體包括：

在所述目標二部圖中刪除權重小於第一預設權重的邊，以得到待聚類二部圖，並對所述待聚類二部圖採用聯通演算法得到至少一個最大連通子圖，以及將每個所述最大連通子圖中的節點對應的待分析用戶確定為一個所述異常群體；或者

在所述目標二部圖中刪除權重小於第一預設權重的邊，以得到待聚類二部圖，並透過社區發現演算法對所述待聚類二部圖中的節點進行劃分，以得到多個節點集合，以及將每個所述節點集合中的節點對應的待分析用戶確定為一個所述異常群體。

可選地，所述根據所述目標二部圖中的邊的權重，以及透過對所述目標二部圖進行圖聚類所得到的所述多個待分析用戶的聚類結果，確定所述待分析用戶中的異常群體包括：

根據所述目標二部圖中的邊的權重計算任意兩個所述待分析用戶之間的權重；

將各所述待分析用戶轉化為節點，並在任意兩個節點之間設置邊，並將任意兩個節點的邊的權重設定為對應的任意兩個所述待分析用戶之間的權重，以構建目標聚類圖；

透過對所述目標聚類圖進行圖聚類所得到的所述多個待分析用戶的聚類結果，確定所述待分析用戶中的異常群體。

可選地，所述透過對所述目標聚類圖進行圖聚類所得到的所述多個待分析用戶的聚類結果，確定所述待分析用戶中的異常群體包括：

在所述目標聚類圖中刪除權重小於第二預設權重的邊，以得到待聚類圖，並對所述待聚類圖採用聯通演算法得到至少一個最大連通子圖，以及將每個所述最大連通子圖中的節點對應的待分析用戶分別確定為一個所述異常群體；或者

在所述目標聚類圖中刪除權重小於第二預設權重的邊，以得到待聚類圖，並透過社區發現演算法對所述待聚類圖進行劃分，以得到多個節點集合，以及將每個所述節點集合對應的待分析用戶分別確定為一個所述異常群體。

另一方面，本說明書一個或多個實施例提供一種異常群體識別裝置，包括：

獲取模組，用於獲取多個待分析用戶中的各所述待分析用戶的特徵值；

確定模組，用於確定各所述待分析用戶的特徵值中的

高頻特徵值和低頻特徵值；

挖掘模組，用於根據各所述待分析用戶的高頻特徵值和預設的頻繁項集挖掘策略挖掘最大頻繁項集，獲取所述最大頻繁項集中的低頻最大頻繁特徵值；

構建模組，用於根據各所述待分析用戶的特徵值中的所述低頻最大頻繁特徵值和所述低頻特徵值構建目標二部圖，並定義所述目標二部圖中的邊的權重；

聚類別模組，用於根據所述目標二部圖中的邊的權重，以及透過對所述目標二部圖進行圖聚類所得到的所述多個待分析用戶的聚類結果，確定所述待分析用戶中的異常群體。

可選地，所述獲取模組包括：

獲取單元，用於獲取所述多個待分析用戶的原始個人資料；

離散化單元，用於對所述多個待分析用戶的原始個人資料進行離散化，以得到各所述待分析用戶的特徵值。

可選地，所述確定模組包括：

第一構建單元，用於根據各所述待分析用戶的特徵值構建第一二部圖，其中，所述第一二部圖包括與各所述待分析用戶對應的節點、與各所述特徵值對應的節點、以及各所述待分析用戶對應的節點與其特徵值對應的節點之間的邊；

第一確定單元，用於在所述第一二部圖中獲取各所述特徵值對應的節點的度，並根據各所述特徵值對應的節點

的度在所述特徵值中確定高頻特徵值和低頻特徵值；

第二確定單元，用於根據所述高頻特徵值和所述低頻特徵值確定各所述待分析用戶的特徵值中的高頻特徵值和低頻特徵值。

可選地，所述挖掘模組包括：

挖掘單元，用於根據各所述待分析用戶的高頻特徵值並結合FP-Growth方法，挖掘支援度滿足預設支援度的頻繁多項集，並在所述頻繁多項集中確定最大頻繁項集；

匹配單元，用於將各所述待分析用戶的特徵值與所述最大頻繁項集中的最大頻繁特徵值進行匹配，以得到各所述待分析用戶的最大頻繁特徵值；

第三確定單元，用於在所述待分析用戶的最大頻繁特徵值中確定低頻最大頻繁特徵值。

可選地，所述第三確定單元包括：

構建子單元，用於根據各所述待分析用戶的最大頻繁特徵值構建第二二部圖，其中，所述第二二部圖包括與各所述待分析用戶對應的節點、與各所述最大頻繁特徵值對應的節點、以及各所述待分析用戶對應的節點與其最大頻繁特徵值對應的節點之間的邊；

確定子單元，用於在所述第二二部圖中獲取各所述最大頻繁特徵值對應的節點的度，並根據各所述最大頻繁特徵值對應的節點的度在所述最大頻繁特徵值中確定低頻最大頻繁特徵值。

可選地，所述聚類別模組包括：

第一聚類單元，用於在所述目標二部圖中刪除權重小於第一預設權重的邊，以得到待聚類二部圖，並對所述待聚類二部圖採用聯通演算法得到至少一個最大連通子圖，以及將每個所述最大連通子圖中的節點對應的待分析用戶確定為一個所述異常群體；或者

第二聚類單元，用於在所述目標二部圖中刪除權重小於第一預設權重的邊，以得到待聚類二部圖，並透過社區發現演算法對所述待聚類二部圖中的節點進行劃分，以得到多個節點集合，以及將每個所述節點集合中的節點對應的待分析用戶確定為一個所述異常群體。

可選地，所述聚類別模組包括：

計算單元，用於根據所述目標二部圖中的邊的權重計算任意兩個所述待分析用戶之間的權重；

第二構建單元，用於將各所述待分析用戶轉化為節點，並在任意兩個節點之間設置邊，並將任意兩個節點的邊的權重設定為對應的任意兩個所述待分析用戶之間的權重，以構建目標聚類圖；

第三聚類單元，用於透過對所述目標聚類圖進行圖聚類所得到的所述多個待分析用戶的聚類結果，確定所述待分析用戶中的異常群體。

可選地，所述第三聚類單元包括：

第一聚類子單元，用於在所述目標聚類圖中刪除權重小於第二預設權重的邊，以得到待聚類圖，並對所述待聚類圖採用聯通演算法得到至少一個最大連通子圖，以及將

每個所述最大連通子圖中的節點對應的待分析用戶分別確定為一個所述異常群體；或者

第二聚類子單元，用於在所述目標聚類圖中刪除權重小於第二預設權重的邊，以得到待聚類圖，並透過社區發現演算法對所述待聚類圖進行劃分，以得到多個節點集合，以及將每個所述節點集合對應的待分析用戶分別確定為一個所述異常群體。

再一方面，本說明書一個或多個實施例提供一種異常群體識別設備，包括：

處理器；以及

被安排成儲存電腦可執行指令的記憶體，所述電腦可執行指令在被執行時使所述處理器：

獲取多個待分析用戶中的各所述待分析用戶的特徵值；

確定各所述待分析用戶的特徵值中的高頻特徵值和低頻特徵值；

根據各所述待分析用戶的高頻特徵值和預設的頻繁項集挖掘策略挖掘最大頻繁項集，獲取所述最大頻繁項集中的低頻最大頻繁特徵值；

根據各所述待分析用戶的特徵值中的所述低頻最大頻繁特徵值和所述低頻特徵值構建目標二部圖，並定義所述目標二部圖中的邊的權重；

根據所述目標二部圖中的邊的權重，以及透過對所述目標二部圖進行圖聚類所得到的所述多個待分析用戶的聚

類結果，確定所述待分析用戶中的異常群體。

再一方面，本說明書一個或多個實施例提供一種儲存媒體，用於儲存電腦可執行指令，所述電腦可執行指令在被執行時實現以下流程：

獲取多個待分析用戶中的各所述待分析用戶的特徵值；

確定各所述待分析用戶的特徵值中的高頻特徵值和低頻特徵值；

根據各所述待分析用戶的高頻特徵值和預設的頻繁項集挖掘策略挖掘最大頻繁項集，獲取所述最大頻繁項集中的低頻最大頻繁特徵值；

根據各所述待分析用戶的特徵值中的所述低頻最大頻繁特徵值和所述低頻特徵值構建目標二部圖，並定義所述目標二部圖中的邊的權重；

根據所述目標二部圖中的邊的權重，以及透過對所述目標二部圖進行圖聚類所得到的所述多個待分析用戶的聚類結果，確定所述待分析用戶中的異常群體。

採用本說明書一個或多個實施例的技術方案，透過確定各待分析用戶的特徵值中的高頻特徵值和低頻特徵值，並透過對各待分析用戶的高頻特徵值進行預設的頻繁項集挖掘策略挖掘最大頻繁項集，獲取最大頻繁項集中的低頻最大頻繁特徵值，以及根據各待分析用戶的低頻特徵值和低頻最大頻繁特徵值構建目標二部圖，並設定目標二部圖中的邊的權重，以根據目標二部圖中的邊的權重以及對目

標二部圖進行聚類，以確定待分析用戶中的異常群體。一方面，透過對各待分析用戶的高頻特徵值進行預設的頻繁項集挖掘策略挖掘最大頻繁項集，並獲取最大頻繁項集中的低頻最大頻繁特徵值，以挖掘待分析用戶的行為序列，進而使得異常群體的識別更加準確；另一方面，僅透過獲取各待分析用戶的低頻特徵值和低頻最大頻繁特徵值，並根據各待分析用戶的低頻特徵值和低頻最大頻繁特徵值構建目標二部圖，並定義目標二部圖中的邊的權重，以及根據目標二部圖中的邊的權重並對目標二部圖進行圖聚類，以得到異常群體，步驟簡單，且易於執行。

【圖式簡單說明】

為了更清楚地說明本說明書一個或多個實施例或現有技術中的技術方案，下面將對實施例或現有技術描述中所需要使用的圖式作簡單地介紹，顯而易見地，下面描述中的圖式僅僅是本說明書一個或多個實施例中記載的一些實施例，對於本發明所屬技術領域中具有通常知識者來講，在不付出創造性勞動性的前提下，還可以根據這些圖式獲得其他的圖式。

圖1為本發明實施例提供的異常群體識別方法的流程示意圖；

圖2為本發明實施例提供的確定各待分析用戶的特徵值中的高頻特徵值和低頻特徵值的流程示意圖；

圖3為本發明實施例提供的第一二部圖的示意圖；

圖4為本發明實施例提供的獲取低頻最大頻繁特徵值的流程示意圖一；

圖5為本發明實施例提供的獲取低頻最大頻繁特徵值的流程示意圖二；

圖6為本發明實施例提供的確定異常群體的流程示意圖；

圖7為本發明實施例提供的異常群體識別裝置的組成示意圖；

圖8為本發明實施例提供的異常群體識別設備的結構示意圖。

【實施方式】

本說明書一個或多個實施例提供一種異常群體識別方法及裝置，用以解決現有技術中異常群體識別準確率較低的問題。

為了使本技術領域的人員更好地理解本說明書一個或多個實施例中的技術方案，下面將結合本說明書一個或多個實施例中的圖式，對本說明書一個或多個實施例中的技術方案進行清楚、完整地描述，顯然，所描述的實施例僅僅是本說明書一部分實施例，而不是全部的實施例。基於本說明書一個或多個實施例，本領域普通技術人員在沒有作出創造性勞動前提下所獲得的所有其他實施例，都應當屬於本說明書一個或多個實施例保護的範圍。

圖1為本發明實施例提供的異常群體識別方法的流程

示意圖，該方法的執行主體例如可以為終端設備或伺服器，其中，終端設備例如可以為個人電腦等，伺服器例如可以為獨立的一個伺服器，也可以是由多個伺服器組成的伺服器集群，本示例性實施例對此不做特殊限定。如圖1所示，該方法可以包括以下步驟：

步驟S102、獲取多個待分析用戶中的各待分析用戶的特徵值。

在本發明實施例中，可以首先獲取多個待分析用戶的原始個人資料，然後，對多個待分析用戶的原始個人資料進行離散化，以得到各待分析用戶的特徵值。其中，獲取多個待分析用戶的原始個人資料包括：可以透過一獲取模組獲取各待分析用戶的原始個人資料，並將各待分析用戶的原始個人資料進行集合得到多個待分析用戶的原始個人資料。每個待分析用戶的原始個人資料均可以包括個人基本資料、行為資料、設備資料等，本示例性實施例對此不做特殊限定。個人基本資料中可以包括年齡、性別、職業、收入、學歷、籍貫、聯繫方式、帳號等特徵的資料，本示例性實施例對此不做特殊限定。例如，個人基本資料可以包括：女(性別)、18歲(年齡)、本科(學歷)、律師(職業)、陝西(籍貫)。行為資料可以包括多個行為特徵的資料，具體的，行為資料中包括的行為特徵的資料可以根據應用場景的不同進行設定。例如，在保險場景下，行為資料可以包括：2018.10.03號投保(投保時間)、意外險(投保種類)、2019.2.1號出險(出險特徵)等。設備資料例如可以

包括：設備型號、設備歸屬地、使用設備的常用位址、更換設備的頻率等特徵的資料，本示例性實施例對此不做特殊限定。

對多個待分析用戶的原始個人資料進行離散化，以得到各待分析用戶的特徵值可以包括：根據多個待分析用戶的原始個人資料中的各特徵的資料分析各特徵的資料的分佈，再根據各特徵的資料的分佈並結合分箱方式對各特徵的資料進行分箱，並將各特徵的資料分箱後對應的區間確定為對應的各特徵的資料的特徵值，以及根據各特徵的資料的特徵值並結合各待分析用戶的原始個人資料確定各待分析用戶的特徵值。

分箱方式可以根據特徵所屬的性質進行確定，對於連續型的特徵(例如年齡、收入、交易金額等)，可以根據業務經驗和資料分佈確定採用等頻、等寬等分箱方式。對於類別型的特徵(例如，性別、學歷、職業等)，可以根據特徵的具體類別對類別型的特徵的資料進行分箱。對於文字型的特徵(例如位址等)，可以採用將模式一致的文字聚成一類的方式進行分箱。

需要說明的是，可以根據待分析用戶的唯一標識對待分析用戶進行標記，用以區分待分析用戶。唯一標識例如可以為：身分證、軍官證、帳號id等，本示例性實施例對此不做特殊限定。

步驟S104、確定各待分析用戶的特徵值中的高頻特徵值和低頻特徵值。

在本示例性實施例中，可以透過以下兩種方式確定待分析用戶的特徵值中的高頻特徵值和低頻特徵值，其中：

方式一、統計每個特徵值在多個待分析用戶的特徵值中出現的次數，並根據下述確定規則在特徵值中確定高頻特徵值和低頻特徵值，其中，確定規則為：若特徵值在多個待分析用戶的特徵值中出現的次數符合公式$T2_i \geq X_i > T1_i$，則特徵值為低頻特徵值，其中，X_i為第i個特徵值在多個待分析用戶的特徵值中出現的次數，$T2_i$為第i個特徵值對應的第二預設出現次數，$T1_i$為第i個特徵值對應的第一預設出現次數，$T2_i > T1_i$，且$T2_i$和$T1_i$的具體數值可以根據第i個特徵值所屬的特徵進行確定，即特徵不同，對應的$T2_i$和$T1_i$的具體數值也不同；若特徵值在多個待分析用戶的特徵值中出現的次數符合公式$T3_i \geq X_i > T2_i$，則特徵值為高頻特徵值，其中，X_i為第i個特徵值在多個待分析用戶的特徵值中出現的次數，$T2_i$為第i個特徵值對應的第二預設出現次數，$T3_i$為第i個特徵值對應的第三預設出現次數，$T3_i > T2_i$，且$T2_i$和$T3_i$的具體數值可以根據第i個特徵值所屬的特徵進行確定，即特徵不同，對應的$T2_i$和$T3_i$的具體數值也不同。

在確定出高頻特徵值和低頻特徵值後，可以透過將高頻特徵值和低頻特徵分別與各待分析用戶的特徵值進行匹配，以得到各待分析用戶的高頻特徵值和低頻特徵值。例如，高頻特徵值包括：A、B、D，低頻特徵值包括C、E，若待分析用戶的特徵值包括：A、B、C、E，則該待分析

用戶的高頻特徵值包括A、B，該待分析用戶的低頻特徵值包括C、E；若待分析用戶的特徵值包括：A、E、F，則該待分析用戶的高頻特徵值包括A，該待分析用戶的低頻特徵值包括E。

方式二、如圖2所示，可以包括以下步驟：

步驟S202、根據各待分析用戶的特徵值構建第一二部圖，其中，第一二部圖包括與各待分析用戶對應的節點、與各特徵值對應的節點、以及各待分析用戶對應的節點與其特徵值對應的節點之間的邊。

在本發明實施例中，將每個待分析用戶分別轉化為節點，每個待分析用戶僅對應一個節點，並將各待分析用戶的特徵值轉化為節點，每個特徵值僅對應一個節點，即在轉化的過程中，若一個特徵值對應的節點已經存在，則複用該節點，無需再設置與該特徵值對應的節點，其中，與各待分析用戶對應的節點位於第一二部圖的一側，與各特徵值對應的節點位於第一二部圖的另一側，且在與各待分析用戶對應的節點與其特徵值對應的節點之間添加邊。例如，待分析用戶為5個，分別為第一待分析用戶至第五待分析用戶，其中，第一待分析用戶的特徵值包括：A、B、D，第二待分析用戶的特徵值包括：B、C、F，第三待分析用戶的特徵值包括：A、C、D、F，第四待分析用戶的特徵值包括：B、D、F，第五待分析用戶的特徵值包括：C、D、E、F，基於此，構建的第一二部圖如圖3所示，其中，第一待分析用戶對應的節點1、第二待分析用

戶對應的節點2、第三待分析用戶對應的節點3、第四待分析用戶對應的節點4以及第五待分析用戶對應的節點5位於圖3的左側，特徵值A對應的節點、特徵值B對應的節點、特徵值C對應的節點、特徵值D對應的節點、特徵值E對應的節點、特徵值F對應的節點位於圖3的右側，且在各待分析用戶對應的節點和其特徵值對應的節點之間設置邊。

步驟S204、在第一二部圖中獲取各特徵值對應的節點的度，並根據各特徵值對應的節點的度在特徵值中確定高頻特徵值和低頻特徵值。

在本發明實施例中，特徵值對應的節點的度指與特徵值對應的節點連接的邊的數量，例如，在圖3中，特徵值A對應的節點的度為2、特徵值B對應的節點的度為3、特徵值C對應的節點的度為3、特徵值D對應的節點的度為4、特徵值E對應的節點的度為1、特徵值F的度為4。

根據各特徵值對應的節點的度在特徵值中確定高頻特徵值和低頻特徵值的過程可以包括：根據各特徵值並結合下述確定規則確定高頻特徵值和低頻特徵值，其中確定規則可以為：若特徵值對應的節點的度滿足公式$K2_i \geq degree(V_i) > 1$，則特徵值為低頻特徵值，其中，$degree(V_i)$為第i個特徵值$V_i$對應的節點的度，$K2_i$為第i個特徵值$V_i$對應的第一預設度，$K2_i > 1$，且$K2_i$的具體數值可以根據第i個特徵值$V_i$所屬的特徵進行確定，即特徵不同，對應的$K2_i$的具體數值也不同；若特徵值對應的節點的度滿足公式$K1_i \geq degree(V_i) > K2_i$，則特徵值為高頻特徵

值，其中，degree(V_i)為第i個特徵值V_i對應的節點的度，$K2_i$為第i個特徵值V_i對應的第一預設度，$K1_i$為第i各特徵值V_i對應的第二預設度，$K1_i > K2_i$，且$K2_i$和$K1_i$的具體數值可以根據第i個特徵值V_i所屬的特徵進行確定，即特徵不同，對應的$K2_i$和$K1_i$的具體數值也不同。

例如，如圖3所示，若$K2_i$為2，$K1_i$為3，則特徵值A為低頻特徵值，特徵值B、特徵值C為高頻特徵值。

步驟S206、根據高頻特徵值和低頻特徵值確定各待分析用戶的特徵值中的高頻特徵值和低頻特徵值。

在本發明實施例中，將高頻特徵值分別與各待分析用戶的特徵值進行匹配，並將各待分析用戶中的與高頻特徵值匹配成功的特徵值確定為對應的各待分析用戶的高頻特徵值；將低頻特徵值分別與各待分析用戶中的特徵值進行匹配，並將各待分析用戶中的與低頻特徵值匹配成功的特徵值確定為對應的各待分析用戶的低頻特徵值。例如，如圖3所示，若$K2_i$為2，$K1_i$為3，則特徵值A為低頻特徵值，特徵值B、特徵值C為高頻特徵值。基於此，第一待分析用戶的低頻特徵值包括特徵值A、第一待分析用戶的高頻特徵值包括特徵值B，第二待分析用戶沒有低頻特徵值，第二待分析用戶的高頻特徵值包括：特徵值B、特徵值C，第三待分析用戶的低頻特徵值包括特徵值A，第三待分析用戶的高頻特徵值包括特徵值C，第四待分析用戶沒有低頻特徵值，第四待分析用戶的高頻特徵值包括特徵值B，第五待分析用戶沒有低頻特徵值，第五待分析用戶的

高頻特徵值包括特徵值C。

步驟S106、根據各待分析用戶的高頻特徵值和預設的頻繁項集挖掘策略挖掘最大頻繁項集，獲取最大頻繁項集中的低頻最大頻繁特徵值。

在本發明實施例中，預設的頻繁項集挖掘策略例如可以為Apriori(挖掘關聯規則的頻繁項集)策略，還可以為FP-Growth等，本示例性實施例對此不做特殊限定。下面，以預設的頻繁項集挖掘策略為FP-Growth為例，對上述過程進行說明，其中，如圖4所示，可以包括以下步驟：

步驟S402、根據各待分析用戶的高頻特徵值並結合FP-Growth方法，挖掘支援度滿足預設支援度的頻繁多項集，並在頻繁多項集中確定最大頻繁項集。

在本發明實施例中，支持度為高頻特徵值在多個待分析用戶中的出現次數，預設支援度的具體數值可以自行設定，例如可以為1、也可以為2等，本示例性實施例對此不做特殊限定。頻繁多項集指至少包括兩個高頻特徵值的集合。支持度滿足預設支持度的頻繁多項集指頻繁多項集中的每個高頻特徵值的支持度均大於預設支援度。

具體的挖掘頻繁多項集的過程包括：定義預設支援度，掃描各待分析用戶的高頻特徵值，以得到每個高頻特徵值在多個待分析用戶中的出現次數(即支持度)，並在各待分析用戶的高頻特徵值中篩除支持度小於預設支持度的高頻特徵值，以及根據各待分析用戶中剩餘的高頻特徵值

構建FP樹，並在FP樹中挖掘頻繁多項集。在頻繁多項集中獲取無超集合條件的頻繁多項集，並將頻繁多項集中的無超集合條件的頻繁多項集確定為最大頻繁項集。需要說明的是，每個最大頻繁項集中包括多個高頻特徵值，此處，將最大頻繁項集中包括的高頻特徵值命名為最大頻繁特徵值，即每個最大頻繁項集中包括多個最大頻繁特徵值。

步驟S404、將各待分析用戶的特徵值與最大頻繁項集中的最大頻繁特徵值進行匹配，以得到各待分析用戶的最大頻繁特徵值。

在本發明實施例中，將各待分析用戶的特徵值與最大頻繁項集中的最大頻繁特徵值進行匹配，並將各待分析用戶中與最大頻繁項集中的最大頻繁特徵值匹配成功的特徵值確定為對應的各待分析用戶的最大頻繁特徵值。

步驟S406、在待分析用戶的最大頻繁特徵值中確定低頻最大頻繁特徵值。

在本發明實施例中，可以透過以下兩種方式確定低頻最大頻繁特徵值，其中：

方式一、根據各待分析用戶的最大頻繁特徵值統計各最大頻繁特徵值在多個待分析用戶中的出現次數，並根據各最大頻繁特徵值在多個待分析用戶中的出現次數並結合下述確定規則在最大頻繁特徵值中確定低頻最大頻繁特徵值，其中，確定規則為：若最大頻繁特徵值在多個待分析用戶中的出現次數符合公式$P2_i \geq S_i$，則最大頻繁特徵值為低頻最大頻繁特徵值，其中，$P2_i$為第i個最大頻繁特徵值

對應的預設出現次數，且$P2_i$的具體數值可以根據第i個最大頻繁特徵值所屬的特徵進行確定，即特徵不同，對應的$P2_i$的具體數值也不同，S_i為第i個最大頻繁特徵值在多個待分析用戶中的出現次數。

方式二、如圖5所示，可以包括以下步驟：

步驟S502、根據各待分析用戶的最大頻繁特徵值構建第二二部圖，其中，第二二部圖包括與各待分析用戶對應的節點、與各最大頻繁特徵值對應的節點、以及各待分析用戶對應的節點與其最大頻繁特徵值對應的節點之間的邊。

在本發明實施例中，將每個待分析用戶分別轉化為節點，每個待分析用戶僅對應一個節點，並將各待分析用戶的最大頻繁特徵值轉化為節點，每個最大頻繁特徵值僅對應一個節點，其中，與各待分析用戶對應的節點位於第二二部圖的一側，與各最大頻繁特徵值對應的節點位於第二二部圖的另一側，且在各待分析用戶對應的節點與其最大頻繁特徵值對應的節點之間添加邊，以完成對第二二部圖的構建。

步驟S504、在第二二部圖中獲取各最大頻繁特徵值對應的節點的度，並根據各最大頻繁特徵值對應的節點的度在最大頻繁特徵值中確定低頻最大頻繁特徵值。

在本發明實施例中，最大頻繁特徵值對應的節點的度為二部圖中與該最大頻繁特徵值對應的節點相連的邊的數量。確定低頻最大頻繁特徵值的過程可以包括：根據各最

大頻繁特徵值對應的節點的度並結合下述確定規則確定低頻最大頻繁特徵值，其中確定規則可以為：若最大頻繁特徵值對應的節點的度滿足公式$L2_i \geq degree(V_i)$，則最大頻繁特徵值為低頻最大頻繁特徵值，其中，$degree(V_i)$為第i個最大頻繁特徵值對應的節點的度，$L2_i$第i個最大頻繁特徵值V_i對應的預設度，且$L2_i$的具體數值可以根據第i個最大頻繁特徵值V_i所屬的特徵進行確定，即特徵不同，對應的$L2_i$的具體數值也不同。

步驟S108、根據各待分析用戶的特徵值中的低頻最大頻繁特徵值和低頻特徵值構建目標二部圖，並定義目標二部圖中的邊的權重。

在本發明實施例中，將低頻最大頻繁特徵值與各待分析用戶中的特徵值進行匹配，並將各待分析用戶中與低頻最大頻繁特徵值匹配成功的特徵值確定為對應的各待分析用戶的低頻最大頻繁特徵值。根據各待分析用戶的低頻最大頻繁特徵值以及步驟S104中獲取的各待分析用戶的低頻特徵值構建目標二部圖的過程可以包括：將各待分析用戶分別轉化為節點，並將各低頻特徵值轉化為節點，將各低頻最大頻繁特徵值轉化為節點，以及在各待分析用戶對應的節點與其低頻特徵值對應的節點之間添加邊，並在各待分析用戶對應的節點與其低頻最大頻繁特徵值對應的節點之間添加邊，以完成對目標二部圖的構建。

定義目標二部圖中的邊的權重可以包括：定義目標二部圖中各待分析用戶對應的節點與其低頻特徵值對應的節

點之間的邊的權重，以及定義目標二部圖中各待分析用戶對應的節點與其低頻最大頻繁特徵值對應的節點之間的邊的權重。其中，定義目標二部圖中各待分析用戶對應的節點與其低頻特徵值對應的節點之間的邊的權重可以包括：根據各低頻特徵值所屬的特徵確定各低頻特徵值的權重，具體地，低頻特徵值的權重越高，同時包括該低頻特徵值的待分析用戶為一個異常群體的概率越高，低頻特徵值的權重越低，同時包括該低頻特徵值的待分析用戶為一個異常群體的概率越低。在確定各低頻特徵值的權重後，將與各低頻特徵值對應的節點連接的邊的權重均設定為對應的各低頻特徵值的權重。例如，若低頻特徵值包括頻繁出險(出險特徵對應的特徵值)、無業(職業特徵對應的特徵值)，且頻繁出險的權重為0.5、無業的權重為0.1，則，與頻繁出險對應的節點連接的邊的權重均設定為0.5，與無業對應的節點連接的邊的權重均設定為0.1。同理，定義目標二部圖中各待分析用戶對應的節點與其低頻最大頻繁特徵值對應的節點之間的邊的權重可以包括：根據各低頻最大頻繁特徵值所屬的特徵確定各低頻最大頻繁特徵值的權重，具體地，低頻最大頻繁特徵值的權重越高，同時包括該低頻最大頻繁特徵值的待分析用戶為一個異常群體的概率越高，低頻最大頻繁特徵值的權重越低，同時包括該低頻最大頻繁特徵值的待分析用戶為一個異常群體的概率越低。將與各低頻最大頻繁特徵值對應的節點連接的邊的權重設定為對應的各低頻最大頻繁特徵值的權重。

步驟S110、根據目標二部圖中的邊的權重，以及透過對目標二部圖進行圖聚類所得到的多個待分析用戶的聚類結果，確定待分析用戶中的異常群體。

在本發明實施例中，可以透過以下兩種方式確定待分析用戶中的異常群體，其中：

方式一、在目標二部圖中刪除權重小於第一預設權重的邊，以得到待聚類二部圖，並對待聚類二部圖採用聯通演算法得到至少一個最大連通子圖，以及將每個最大連通子圖中的節點對應的待分析用戶確定為一個異常群體。

在本發明實施例中，第一預設權重的具體數值可以自行設定，本示例性實施例對此不做特殊限定。將目標二部圖中的每個邊的權重依次與第一預設權重進行比較，若邊的權重小於第一預設權重，則在目標二部圖中刪除該邊，若邊的權重不小於第一預設權重，則在目標二部圖中保留該邊，將篩除權重小於預設權重的邊的目標二部圖確定為待聚類二部圖。對待聚類二部圖採用聯通演算法以得到至少一個最大連通子圖，在每個最大連通子圖中篩除與低頻特徵值對應的節點和與低頻最大頻繁特徵值對應的節點，並將每個最大連通子圖中剩餘的節點對應的待分析用戶進行集合，以得到每個最大連通子圖對應的待分析用戶集合，以及將每個最大連通子圖對應的待分析用戶集合分別確定為一個異常群體。

方式二、在目標二部圖中刪除權重小於第一預設權重的邊，以得到待聚類二部圖，並透過社區發現演算法對待

聚類二部圖中的節點進行劃分，以得到多個節點集合，以及將每個節點集合中的節點對應的待分析用戶確定為一個異常群體。

在本發明實施例中，由於在二部圖中刪除權重小於第一預設權重的邊，以得到待聚類二部圖的原理與上述方式一中的原理相同，因此此處不在贅述。社區發現演算法例如可以為louvain演算法等，本示例性實施例對此不做特殊限定。在透過社區發現演算法對待聚類二部圖中的節點進行劃分得到多個節點集合後，首先在每個節點集合中篩除與低頻特徵值對應的節點和低頻最大頻繁特徵值對應的節點，並分別將每個節點集合中剩餘的節點對應的待分析用戶進行集合，以得到每個節點集合對應的待分析用戶集合，並將每個節點集合對應的待分析用戶集合分別確定為一個異常群體。

進一步地，在得到異常群體之後，為了進一步對異常群體進行驗證，進而進一步的提高異常群體識別的準確度，可以獲取每個異常群體中的待分析用戶的總數量，並在異常群體中篩除待分析用戶的總數量少於預設數量的異常群體，並將剩餘的異常群體確定為最終識別出的異常群體；還可以計算每個異常群體對應的最大連通子圖的模組度，並將每個異常群體對應的最大連通子圖的模組度確定為對應的異常群體的模組度，以及在異常群體中篩除模組度小於預設模組度的異常群體，將剩餘的異常群體確定為最終識別出的異常群體。需要說明的是，上述兩種驗證方

式僅為示例性的，並不用於限定本發明，其還可以透過分析異常群體中的每個待分析用戶的業務特徵對異常群體進行驗證。

為了更加準確的對待分析用戶進行聚類，以得到更加準確的異常群體，如圖6所示，根據目標二部圖中的邊的權重，以及透過對目標二部圖進行圖聚類所得到的多個待分析用戶的聚類結果，確定待分析用戶中的異常群體可以包括以下步驟：

步驟S602、根據目標二部圖中的邊的權重計算任意兩個待分析用戶之間的權重。

在本發明實施例中，在目標二部圖中獲取與任意兩個待分析用戶對應的節點共同連接的與低頻特徵值對應的節點和與低頻最大頻繁特徵值對應的節點，並將與任意兩個待分析用戶對應的節點共同連接的與低頻特徵值對應的節點和與低頻最大頻繁特徵值對應的節點確定為目標節點；根據任意兩個待分析用戶中的任何一個待分析用戶對應的節點與每個目標節點之間的邊的權重並結合下述公式計算任意兩個待分析用戶之間的權重，上述公式為：

$$weight(e)=1-\prod_{i=1}^{j}(1-w(item_i))$$

其中，weight(e)為任意兩個待分析用戶之間的權重，j為目標節點的總數量，$w(item_i)$為第i個目標節點$item_i$與任意兩個待分析用戶中的任意一個待分析用戶對應的節點之間的邊的權重。

步驟S604、將各待分析用戶轉化為節點，並在任意兩個節點之間設置邊，並將任意兩個節點的邊的權重設定為對應的任意兩個待分析用戶之間的權重，以構建目標聚類圖。

在本發明實施例中，將各待分析用戶轉化為節點，即一個待分析用戶僅對應一個節點，並在任意兩個節點之間設置邊，以及將任意兩個待分析用戶之間的權重設定為該任意兩個待分析用戶對應的兩個節點之間的邊的權重，以完成目標聚類圖的構建。由上可知，透過步驟S602和步驟S604將包括待分析用戶對應的節點和低頻特徵值對應的節點以及低頻最大頻繁特徵值對應的節點的目標二部圖轉化為僅包括待分析用戶對應的節點的目標聚類圖。

步驟S606、透過對目標聚類圖進行圖聚類所得到的多個待分析用戶的聚類結果，確定待分析用戶中的異常群體。

在本發明實施例中，可以透過以下兩種方式確定異常群體，其中：

方式一、在目標聚類圖中刪除權重小於第二預設權重的邊，以得到待聚類圖，並對待聚類圖採用聯通演算法得到至少一個最大連通子圖，以及將每個最大連通子圖中的節點對應的待分析用戶分別確定為一個異常群體。

在本發明實施例中，第二預設權重的具體數值可以自行設定，本示例性實施例對此不做特殊限定。將目標聚類圖中的每個邊的權重分別與第二預設權重進行比較，並在

目標聚類圖中刪除權重小於第二預設權重的邊，以將目標聚類圖轉化為待聚類圖。將每個最大連通子圖中的節點對應的待分析用戶進行集合，以得到每個最大連通子圖對應的待分析用戶集合，並將每個最大連通子圖對應的待分析用戶集合分別確定為一個異常群體。

方式二、在目標聚類圖中刪除權重小於第二預設權重的邊，以得到待聚類圖，並透過社區發現演算法對待聚類圖進行劃分，以得到多個節點集合，以及將每個節點集合對應的待分析用戶分別確定為一個異常群體。

在申請實施例中，第二預設權重已經在上文中進行了說明，因此此處不在贅述。將目標聚類圖中的每個邊的權重分別與第二預設權重進行比較，並在目標聚類圖中刪除權重小於第二預設權重的邊，以將目標聚類圖轉化為待聚類圖。社區發現演算法例如可以為louvain演算法等，本示例性實施例對此不做特殊限定。在透過社區發現演算法對待聚類圖中的節點進行劃分得到多個節點集合後，分別將每個節點集合中的節點對應的待分析用戶進行集合，以得到每個節點集合對應的待分析用戶集合，並將每個節點集合對應的待分析用戶集合分別確定為一個異常群體。

由上可知，透過根據目標二部圖中的邊的權重計算任意兩個待分析用戶之間的權重，並根據任意兩個待分析用戶之前的權重構建目標聚類圖，以將目標二部圖轉化為目標聚類圖，使得目標聚類圖更加準確且更加直觀的反應待分析用戶之間的關係，進而使得根據目標聚類圖得到的異

常群體更加準確。

需要說明的是，上述兩種確定異常群體的方式進行示例性的，並不用於限定本發明。

進一步地，在得到異常群體之後，為了進一步對異常群體進行驗證，進而進一步的提高異常群體識別的準確度，可以獲取每個異常群體中的待分析用戶的總數量，並在異常群體中篩除待分析用戶的總數量少於預設數量的異常群體，並將剩餘的異常群體確定為最終識別出的異常群體；還可以計算每個異常群體對應的最大連通子圖的模組度，並將每個異常群體對應的最大連通子圖的模組度確定為對應的異常群體的模組度，以及在異常群體中篩除模組度小於預設模組度的異常群體，將剩餘的異常群體確定為最終識別出的異常群體。需要說明的是，上述兩種驗證方式僅為示例性的，並不用於限定本發明，其還可以透過分析異常群體中的每個待分析用戶的業務特徵對異常群體進行驗證。

綜上所述，透過對各待分析用戶的高頻特徵值進行預設的頻繁項集挖掘策略挖掘最大頻繁項集，並獲取最大頻繁項集中的低頻最大頻繁特徵值，以挖掘待分析用戶的行為序列，進而使得異常群體的識別更加準確；此外，僅透過獲取各待分析用戶的低頻特徵值和低頻最大頻繁特徵值，並根據各待分析用戶的低頻特徵值和低頻最大頻繁特徵值構建目標二部圖，並定義目標二部圖中的邊的權重，以及根據目標二部圖中的邊的權重並對目標二部圖進行圖

聚類，以得到異常群體，步驟簡單，且易於執行。

對應上述異常群體識別方法，基於相同的技術構思，本發明實施例還提供了一種異常群體識別裝置，圖7為本發明實施例提供的異常群體識別裝置700的組成示意圖，該裝置用於執行上述異常群體識別方法，如圖7所示，該裝置700可以包括：獲取模組701、確定模組702、挖掘模組703、構建模組704、聚類別模組705，其中：

獲取模組701，用於獲取多個待分析用戶中的各所述待分析用戶的特徵值；

確定模組702，用於確定各所述待分析用戶的特徵值中的高頻特徵值和低頻特徵值；

挖掘模組703，用於根據各所述待分析用戶的高頻特徵值和預設的頻繁項集挖掘策略挖掘最大頻繁項集，獲取所述最大頻繁項集中的低頻最大頻繁特徵值；

構建模組704，用於根據各所述待分析用戶的特徵值中的所述低頻最大頻繁特徵值和所述低頻特徵值構建目標二部圖，並定義所述目標二部圖中的邊的權重；

聚類別模組705，用於根據所述目標二部圖中的邊的權重，以及透過對所述目標二部圖進行圖聚類所得到的所述多個待分析用戶的聚類結果，確定所述待分析用戶中的異常群體。

可選地，所述獲取模組701可以包括：

獲取單元，用於獲取所述多個待分析用戶的原始個人資料；

離散化單元，用於對所述多個待分析用戶的原始個人資料進行離散化，以得到各所述待分析用戶的特徵值。

可選地，所述確定模組702可以包括：

第一構建單元，用於根據各所述待分析用戶的特徵值構建第一二部圖，其中，所述第一二部圖包括與各所述待分析用戶對應的節點、與各所述特徵值對應的節點、以及各所述待分析用戶對應的節點與其特徵值對應的節點之間的邊；

第一確定單元，用於在所述第一二部圖中獲取各所述特徵值對應的節點的度，並根據各所述特徵值對應的節點的度在所述特徵值中確定高頻特徵值和低頻特徵值；

第二確定單元，用於根據所述高頻特徵值和所述低頻特徵值確定各所述待分析用戶的特徵值中的高頻特徵值和低頻特徵值。

可選地，所述挖掘模組703可以包括：

挖掘單元，用於根據各所述待分析用戶的高頻特徵值並結合FP-Growth方法，挖掘支援度滿足預設支援度的頻繁多項集，並在所述頻繁多項集中確定最大頻繁項集；

匹配單元，用於將各所述待分析用戶的特徵值與所述最大頻繁項集中的最大頻繁特徵值進行匹配，以得到各所述待分析用戶的最大頻繁特徵值；

第三確定單元，用於在所述待分析用戶的最大頻繁特徵值中確定低頻最大頻繁特徵值。

可選地，所述第三確定單元可以包括：

構建子單元，用於根據各所述待分析用戶的最大頻繁特徵值構建第二二部圖，其中，所述第二二部圖包括與各所述待分析用戶對應的節點、與各所述最大頻繁特徵值對應的節點、以及各所述待分析用戶對應的節點與其最大頻繁特徵值對應的節點之間的邊；

確定子單元，用於在所述第二二部圖中獲取各所述最大頻繁特徵值對應的節點的度，並根據各所述最大頻繁特徵值對應的節點的度在所述最大頻繁特徵值中確定低頻最大頻繁特徵值。

可選地，所述聚類別模組705可以包括：

第一聚類單元，用於在所述目標二部圖中刪除權重小於第一預設權重的邊，以得到待聚類二部圖，並對所述待聚類二部圖採用聯通演算法得到至少一個最大連通子圖，以及將每個所述最大連通子圖中的節點對應的待分析用戶確定為一個所述異常群體；或者

第二聚類單元，用於在所述目標二部圖中刪除權重小於第一預設權重的邊，以得到待聚類二部圖，並透過社區發現演算法對所述待聚類二部圖中的節點進行劃分，以得到多個節點集合，以及將每個所述節點集合中的節點對應的待分析用戶確定為一個所述異常群體。

可選地，所述聚類別模組705可以包括：

計算單元，用於根據所述目標二部圖中的邊的權重計算任意兩個所述待分析用戶之間的權重；

第二構建單元，用於將各所述待分析用戶轉化為節

點，並在任意兩個節點之間設置邊，並將任意兩個節點的邊的權重設定為對應的任意兩個所述待分析用戶之間的權重，以構建目標聚類圖；

第三聚類單元，用於透過對所述目標聚類圖進行圖聚類所得到的所述多個待分析用戶的聚類結果，確定所述待分析用戶中的異常群體。

可選地，所述第三聚類單元可以包括：

第一聚類子單元，用於在所述目標聚類圖中刪除權重小於第二預設權重的邊，以得到待聚類圖，並對所述待聚類圖採用聯通演算法得到至少一個最大連通子圖，以及將每個所述最大連通子圖中的節點對應的待分析用戶分別確定為一個所述異常群體；或者

第二聚類子單元，用於在所述目標聚類圖中刪除權重小於第二預設權重的邊，以得到待聚類圖，並透過社區發現演算法對所述待聚類圖進行劃分，以得到多個節點集合，以及將每個所述節點集合對應的待分析用戶分別確定為一個所述異常群體。

本發明實施例中的異常群體識別裝置，透過對各待分析用戶的高頻特徵值進行預設的頻繁項集挖掘策略挖掘最大頻繁項集，並獲取最大頻繁項集中的低頻最大頻繁特徵值，以挖掘待分析用戶的行為序列，進而使得異常群體的識別更加準確；此外，僅透過獲取各待分析用戶的低頻特徵值和低頻最大頻繁特徵值，並根據各待分析用戶的低頻特徵值和低頻最大頻繁特徵值構建目標二部圖，並定義目

標二部圖中的邊的權重，以及根據目標二部圖中的邊的權重並對目標二部圖進行圖聚類，以得到異常群體，步驟簡單，且易於執行。

應上述異常群體識別方法，基於相同的技術構思，本發明實施例還提供了一種異常群體識別設備，圖8為本發明實施例提供的異常群體識別設備的結構示意圖，該設備用於執行上述的異常群體識別方法。

如圖8所示，異常群體識別設備可因配置或性能不同而產生比較大的差異，可以包括一個或一個以上的處理器801和記憶體802，記憶體802中可以儲存有一個或一個以上儲存應用程式或資料。其中，記憶體802可以是短暫儲存或持久儲存。儲存在記憶體802的應用程式可以包括一個或一個以上模組(圖示未顯示)，每個模組可以包括對異常群體識別設備中的一系列電腦可執行指令。更進一步地，處理器801可以設定為與記憶體802通訊，在異常群體識別設備上執行記憶體802中的一系列電腦可執行指令。異常群體識別設備還可以包括一個或一個以上電源803、一個或一個以上有線或無線網路介面804、一個或一個以上輸入輸出介面805、一個或一個以上鍵盤806等。

在一個具體的實施例中，異常群體識別設備包括有記憶體，以及一個或一個以上的程式，其中，一個或者一個以上程式儲存於記憶體中，且一個或者一個以上程式可以包括一個或一個以上模組，且每個模組可以包括對異常群體識別設備中的一系列電腦可執行指令，且經配置以由一

個或者一個以上處理器執行該一個或者一個以上套裝程式含用於進行以下電腦可執行指令：

獲取多個待分析用戶中的各所述待分析用戶的特徵值；

確定各所述待分析用戶的特徵值中的高頻特徵值和低頻特徵值；

根據各所述待分析用戶的高頻特徵值和預設的頻繁項集挖掘策略挖掘最大頻繁項集，獲取所述最大頻繁項集中的低頻最大頻繁特徵值；

根據各所述待分析用戶的特徵值中的所述低頻最大頻繁特徵值和所述低頻特徵值構建目標二部圖，並定義所述目標二部圖中的邊的權重；

根據所述目標二部圖中的邊的權重，以及透過對所述目標二部圖進行圖聚類所得到的所述多個待分析用戶的聚類結果，確定所述待分析用戶中的異常群體。

可選地，電腦可執行指令在被執行時，所述獲取多個待分析用戶中的各所述待分析用戶的特徵值包括：

獲取所述多個待分析用戶的原始個人資料；

對所述多個待分析用戶的原始個人資料進行離散化，以得到各所述待分析用戶的特徵值。

可選地，電腦可執行指令在被執行時，所述確定各所述待分析用戶的特徵值中的高頻特徵值和低頻特徵值包括：

根據各所述待分析用戶的特徵值構建第一二部圖，其

中，所述第一二部圖包括與各所述待分析用戶對應的節點、與各所述特徵值對應的節點、以及各所述待分析用戶對應的節點與其特徵值對應的節點之間的邊；

在所述第一二部圖中獲取各所述特徵值對應的節點的度，並根據各所述特徵值對應的節點的度在所述特徵值中確定高頻特徵值和低頻特徵值；

根據所述高頻特徵值和所述低頻特徵值確定各所述待分析用戶的特徵值中的高頻特徵值和低頻特徵值。

可選地，電腦可執行指令在被執行時，所述根據各所述待分析用戶的高頻特徵值和預設的頻繁項集挖掘策略挖掘最大頻繁項集，獲取所述最大頻繁項集中的低頻最大頻繁特徵值包括：

根據各所述待分析用戶的高頻特徵值並結合FP-Growth方法，挖掘支援度滿足預設支援度的頻繁多項集，並在所述頻繁多項集中確定最大頻繁項集；

將各所述待分析用戶的特徵值與所述最大頻繁項集中的最大頻繁特徵值進行匹配，以得到各所述待分析用戶的最大頻繁特徵值；

在所述待分析用戶的最大頻繁特徵值中確定低頻最大頻繁特徵值。

可選地，電腦可執行指令在被執行時，所述在所述待分析用戶的最大頻繁特徵值中確定低頻最大頻繁特徵值包括：

根據各所述待分析用戶的最大頻繁特徵值構建第二二

部圖，其中，所述第二二部圖包括與各所述待分析用戶對應的節點、與各所述最大頻繁特徵值對應的節點、以及各所述待分析用戶對應的節點與其最大頻繁特徵值對應的節點之間的邊；

在所述第二二部圖中獲取各所述最大頻繁特徵值對應的節點的度，並根據各所述最大頻繁特徵值對應的節點的度在所述最大頻繁特徵值中確定低頻最大頻繁特徵值。

可選地，電腦可執行指令在被執行時，，所述根據所述目標二部圖中的邊的權重，以及透過對所述目標二部圖進行圖聚類所得到的所述多個待分析用戶的聚類結果，確定所述待分析用戶中的異常群體包括：

在所述目標二部圖中刪除權重小於第一預設權重的邊，以得到待聚類二部圖，並對所述待聚類二部圖採用聯通演算法得到至少一個最大連通子圖，以及將每個所述最大連通子圖中的節點對應的待分析用戶確定為一個所述異常群體；或者

在所述目標二部圖中刪除權重小於第一預設權重的邊，以得到待聚類二部圖，並透過社區發現演算法對所述待聚類二部圖中的節點進行劃分，以得到多個節點集合，以及將每個所述節點集合中的節點對應的待分析用戶確定為一個所述異常群體。

可選地，電腦可執行指令在被執行時，所述根據所述目標二部圖中的邊的權重，以及透過對所述目標二部圖進行圖聚類所得到的所述多個待分析用戶的聚類結果，確定

所述待分析用戶中的異常群體包括：

根據所述目標二部圖中的邊的權重計算任意兩個所述待分析用戶之間的權重；

將各所述待分析用戶轉化為節點，並在任意兩個節點之間設置邊，並將任意兩個節點的邊的權重設定為對應的任意兩個所述待分析用戶之間的權重，以構建目標聚類圖；

透過對所述目標聚類圖進行圖聚類所得到的所述多個待分析用戶的聚類結果，確定所述待分析用戶中的異常群體。

可選地，電腦可執行指令在被執行時，所述透過對所述目標聚類圖進行圖聚類所得到的所述多個待分析用戶的聚類結果，確定所述待分析用戶中的異常群體包括：

在所述目標聚類圖中刪除權重小於第二預設權重的邊，以得到待聚類圖，並對所述待聚類圖採用聯通演算法得到至少一個最大連通子圖，以及將每個所述最大連通子圖中的節點對應的待分析用戶分別確定為一個所述異常群體；或者

在所述目標聚類圖中刪除權重小於第二預設權重的邊，以得到待聚類圖，並透過社區發現演算法對所述待聚類圖進行劃分，以得到多個節點集合，以及將每個所述節點集合對應的待分析用戶分別確定為一個所述異常群體。

本發明實施例中的異常群體識別設備，透過對各待分析用戶的高頻特徵值進行預設的頻繁項集挖掘策略挖掘最

大頻繁項集，並獲取最大頻繁項集中的低頻最大頻繁特徵值，以挖掘待分析用戶的行為序列，進而使得異常群體的識別更加準確；此外，僅透過獲取各待分析用戶的低頻特徵值和低頻最大頻繁特徵值，並根據各待分析用戶的低頻特徵值和低頻最大頻繁特徵值構建目標二部圖，並定義目標二部圖中的邊的權重，以及根據目標二部圖中的邊的權重並對目標二部圖進行圖聚類，以得到異常群體，步驟簡單，且易於執行。

對應上述異常群體識別方法，基於相同的技術構思，本發明實施例還提供了一種儲存媒體，用於儲存電腦可執行指令，在一個具體的實施例中，該儲存媒體可以為隨身碟、光碟、硬碟等，該儲存媒體儲存的電腦可執行指令在被處理器執行時，能實現以下流程：

獲取多個待分析用戶中的各所述待分析用戶的特徵值；

確定各所述待分析用戶的特徵值中的高頻特徵值和低頻特徵值；

根據各所述待分析用戶的高頻特徵值和預設的頻繁項集挖掘策略挖掘最大頻繁項集，獲取所述最大頻繁項集中的低頻最大頻繁特徵值；

根據各所述待分析用戶的特徵值中的所述低頻最大頻繁特徵值和所述低頻特徵值構建目標二部圖，並定義所述目標二部圖中的邊的權重；

根據所述目標二部圖中的邊的權重，以及透過對所述

目標二部圖進行圖聚類所得到的所述多個待分析用戶的聚類結果，確定所述待分析用戶中的異常群體。

可選地，該儲存媒體儲存的電腦可執行指令在被處理器執行時，所述獲取多個待分析用戶中的各所述待分析用戶的特徵值包括：

獲取所述多個待分析用戶的原始個人資料；

對所述多個待分析用戶的原始個人資料進行離散化，以得到各所述待分析用戶的特徵值。

可選地，該儲存媒體儲存的電腦可執行指令在被處理器執行時，所述確定各所述待分析用戶的特徵值中的高頻特徵值和低頻特徵值包括：

根據各所述待分析用戶的特徵值構建第一二部圖，其中，所述第一二部圖包括與各所述待分析用戶對應的節點、與各所述特徵值對應的節點、以及各所述待分析用戶對應的節點與其特徵值對應的節點之間的邊；

在所述第一二部圖中獲取各所述特徵值對應的節點的度，並根據各所述特徵值對應的節點的度在所述特徵值中確定高頻特徵值和低頻特徵值；

根據所述高頻特徵值和所述低頻特徵值確定各所述待分析用戶的特徵值中的高頻特徵值和低頻特徵值。

可選地，該儲存媒體儲存的電腦可執行指令在被處理器執行時，所述根據各所述待分析用戶的高頻特徵值和預設的頻繁項集挖掘策略挖掘最大頻繁項集，獲取所述最大頻繁項集中的低頻最大頻繁特徵值包括：

根據各所述待分析用戶的高頻特徵值並結合FP-Growth方法，挖掘支援度滿足預設支援度的頻繁多項集，並在所述頻繁多項集中確定最大頻繁項集；

將各所述待分析用戶的特徵值與所述最大頻繁項集中的最大頻繁特徵值進行匹配，以得到各所述待分析用戶的最大頻繁特徵值；

在所述待分析用戶的最大頻繁特徵值中確定低頻最大頻繁特徵值。

可選地，該儲存媒體儲存的電腦可執行指令在被處理器執行時，所述在所述待分析用戶的最大頻繁特徵值中確定低頻最大頻繁特徵值包括：

根據各所述待分析用戶的最大頻繁特徵值構建第二二部圖，其中，所述第二二部圖包括與各所述待分析用戶對應的節點、與各所述最大頻繁特徵值對應的節點、以及各所述待分析用戶對應的節點與其最大頻繁特徵值對應的節點之間的邊；

在所述第二二部圖中獲取各所述最大頻繁特徵值對應的節點的度，並根據各所述最大頻繁特徵值對應的節點的度在所述最大頻繁特徵值中確定低頻最大頻繁特徵值。

可選地，該儲存媒體儲存的電腦可執行指令在被處理器執行時，所述根據所述目標二部圖中的邊的權重，以及透過對所述目標二部圖進行圖聚類所得到的所述多個待分析用戶的聚類結果，確定所述待分析用戶中的異常群體包括：

在所述目標二部圖中刪除權重小於第一預設權重的邊，以得到待聚類二部圖，並對所述待聚類二部圖採用聯通演算法得到至少一個最大連通子圖，以及將每個所述最大連通子圖中的節點對應的待分析用戶確定為一個所述異常群體；或者

在所述目標二部圖中刪除權重小於第一預設權重的邊，以得到待聚類二部圖，並透過社區發現演算法對所述待聚類二部圖中的節點進行劃分，以得到多個節點集合，以及將每個所述節點集合中的節點對應的待分析用戶確定為一個所述異常群體。

可選地，該儲存媒體儲存的電腦可執行指令在被處理器執行時，所述根據所述目標二部圖中的邊的權重，以及透過對所述目標二部圖進行圖聚類所得到的所述多個待分析用戶的聚類結果，確定所述待分析用戶中的異常群體包括：

根據所述目標二部圖中的邊的權重計算任意兩個所述待分析用戶之間的權重；

將各所述待分析用戶轉化為節點，並在任意兩個節點之間設置邊，並將任意兩個節點的邊的權重設定為對應的任意兩個所述待分析用戶之間的權重，以構建目標聚類圖；

透過對所述目標聚類圖進行圖聚類所得到的所述多個待分析用戶的聚類結果，確定所述待分析用戶中的異常群體。

可選地，該儲存媒體儲存的電腦可執行指令在被處理器執行時，所述透過對所述目標聚類圖進行圖聚類所得到的所述多個待分析用戶的聚類結果，確定所述待分析用戶中的異常群體包括：

在所述目標聚類圖中刪除權重小於第二預設權重的邊，以得到待聚類圖，並對所述待聚類圖採用聯通演算法得到至少一個最大連通子圖，以及將每個所述最大連通子圖中的節點對應的待分析用戶分別確定為一個所述異常群體；或者

在所述目標聚類圖中刪除權重小於第二預設權重的邊，以得到待聚類圖，並透過社區發現演算法對所述待聚類圖進行劃分，以得到多個節點集合，以及將每個所述節點集合對應的待分析用戶分別確定為一個所述異常群體。

本發明實施例中的儲存媒體儲存的電腦可執行指令在被處理器執行時，透過對各待分析用戶的高頻特徵值進行預設的頻繁項集挖掘策略挖掘最大頻繁項集，並獲取最大頻繁項集中的低頻最大頻繁特徵值，以挖掘待分析用戶的行為序列，進而使得異常群體的識別更加準確；此外，僅透過獲取各待分析用戶的低頻特徵值和低頻最大頻繁特徵值，並根據各待分析用戶的低頻特徵值和低頻最大頻繁特徵值構建目標二部圖，並定義目標二部圖中的邊的權重，以及根據目標二部圖中的邊的權重並對目標二部圖進行圖聚類，以得到異常群體，步驟簡單，且易於執行。

在20世紀90年代，對於一個技術的改進可以很明顯地

區分是硬體上的改進(例如，對二極體、電晶體、開關等電路結構的改進)還是軟體上的改進(對於方法流程的改進)。然而，隨著技術的發展，當今的很多方法流程的改進已經可以視為硬體電路結構的直接改進。設計人員幾乎都透過將改進的方法流程程式設計到硬體電路中來得到對應的硬體電路結構。因此，不能說一個方法流程的改進就不能用硬體實體模組來實現。例如，可程式設計邏輯器件(Programmable Logic Device，PLD)(例如現場可程式設計閘陣列(Field Programmable Gate Array，FPGA))就是這樣一種積體電路，其邏輯功能由用戶對器件程式設計來確定。由設計人員自行程式設計來把一個數位系統“整合”在一片PLD上，而不需要請晶片製造廠商來設計和製作專用的積體電路晶片。而且，如今，取代手工地製作積體電路晶片，這種程式設計也多半改用“邏輯編譯器(logic Compiler)”軟體來實現，它與程式開發撰寫時所用的軟體編譯器相類似，而要編譯之前的原始代碼也得用特定的程式設計語言來撰寫，此稱之為硬體描述語言(Hardware Description Language，HDL)，而HDL也並非僅有一種，而是有許多種，如ABEL(Advanced Boolean Expression Language)、AHDL(Altera Hardware Description Language)、Confluence、CUPL(Cornell University Programming Language)、HDCal、JHDL(Java Hardware Description Language)、Lava、Lola、MyHDL、PALASM、RHDL(Ruby Hardware Description Language)等，目前最普遍使用

的是 VHDL(Very-High-Speed Integrated Circuit Hardware Description Language)與 Verilog。本領域技術人員也應該清楚，只需要將方法流程用上述幾種硬體描述語言稍作邏輯程式設計並程式設計到積體電路中，就可以很容易得到實現該邏輯方法流程的硬體電路。

控制器可以按任何適當的方式來實現，例如，控制器可以採取例如微處理器或處理器以及儲存可由該(微)處理器執行的電腦可讀程式碼(例如軟體或韌體)的電腦可讀媒體、邏輯閘、開關、專用積體電路(Application Specific Integrated Circuit，ASIC)、可程式設計邏輯控制器和嵌入微控制器的形式，控制器的例子包括但不限於以下微控制器：ARC 625D、Atmel AT91SAM、Microchip PIC18F26K20以及 Silicone Labs C8051F320，記憶體控制器還可以被實現為記憶體的控制邏輯的一部分。本領域技術人員也知道，除了以純電腦可讀程式碼方式實現控制器以外，完全可以透過將方法步驟進行邏輯程式設計來使得控制器以邏輯閘、開關、專用積體電路、可程式設計邏輯控制器和嵌入微控制器等的形式來實現相同功能。因此這種控制器可以被認為是一種硬體部件，而對其內包括的用於實現各種功能的裝置也可以視為硬體部件內的結構。或者甚至，可以將用於實現各種功能的裝置視為既可以是實現方法的軟體模組又可以是硬體部件內的結構。

上述實施例闡明的系統、裝置、模組或單元，具體可以由電腦晶片或實體實現，或者由具有某種功能的產品來

實現。一種典型的實現設備為電腦。具體地，電腦例如可以為個人電腦、膝上型電腦、蜂巢式電話、相機電話、智慧型電話、個人數位助理、媒體播放器、導航設備、電子郵件設備、遊戲控制台、平板電腦、穿戴式設備或者這些設備中的任何設備的組合。

為了描述的方便，描述以上裝置時以功能分為各種單元分別描述。當然，在實施本發明時可以把各單元的功能在同一個或多個軟體和/或硬體中實現。

本領域內的技術人員應明白，本發明的實施例可提供為方法、系統、或電腦程式產品。因此，本發明可採用完全硬體實施例、完全軟體實施例、或結合軟體和硬體方面的實施例的形式。而且，本發明可採用在一個或多個其中包含有電腦可用程式碼的電腦可用儲存媒體(包括但不限於磁碟記憶體、CD-ROM、光學記憶體等)上實施的電腦程式產品的形式。

本發明是參照根據本發明實施例的方法、設備(系統)、和電腦程式產品的流程圖和／或方塊圖來描述的。應理解可由電腦程式指令實現流程圖和／或方塊圖中的每一流程和／或方塊、以及流程圖和／或方塊圖中的流程和／或方塊的結合。可提供這些電腦程式指令到通用電腦、專用電腦、嵌入式處理機或其他可程式設計資料處理設備的處理器以產生一個機器，使得透過電腦或其他可程式設計資料處理設備的處理器執行的指令產生用於實現在流程圖一個流程或多個流程和／或方塊圖一個方塊或多個方塊

中指定的功能的裝置。

這些電腦程式指令也可儲存在能引導電腦或其他可程式設計資料處理設備以特定方式工作的電腦可讀記憶體中，使得儲存在該電腦可讀記憶體中的指令產生包括指令裝置的製造品，該指令裝置實現在流程圖一個流程或多個流程和／或方塊圖一個方塊或多個方塊中指定的功能。

這些電腦程式指令也可裝載到電腦或其他可程式設計資料處理設備上，使得在電腦或其他可程式設計設備上執行一系列操作步驟以產生電腦實現的處理，從而在電腦或其他可程式設計設備上執行的指令提供用於實現在流程圖一個流程或多個流程和／或方塊圖一個方塊或多個方塊中指定的功能的步驟。

在一個典型的配置中，計算設備包括一個或多個處理器(CPU)、輸入/輸出介面、網路介面和記憶體。

記憶體可能包括電腦可讀媒體中的非永久性記憶體，隨機存取記憶體(RAM)和/或非易失性記憶體等形式，如唯讀記憶體(ROM)或快閃記憶體(flash RAM)。記憶體是電腦可讀媒體的示例。

電腦可讀媒體包括永久性和非永久性、可移動和非可移動媒體可以由任何方法或技術來實現資訊儲存。資訊可以是電腦可讀指令、資料結構、程式的模組或其他資料。電腦的儲存媒體的例子包括，但不限於相變記憶體(PRAM)、靜態隨機存取記憶體(SRAM)、動態隨機存取記憶體(DRAM)、其他類型的隨機存取記憶體(RAM)、唯讀

記憶體(ROM)、電可抹除可程式設計唯讀記憶體(EEPROM)、快閃記憶體或其他記憶體技術、唯讀光碟唯讀記憶體(CD-ROM)、數位多功能光碟(DVD)或其他光學儲存、磁盒式磁帶，磁帶磁片儲存或其他磁性儲存設備或任何其他非傳輸媒體，可用於儲存可以被計算設備訪問的資訊。按照本文中的界定，電腦可讀媒體不包括暫態式電腦可讀媒體(transitory media)，如調變的資料訊號和載波。

還需要說明的是，術語“包括”、“包含”或者其任何其他變體意在涵蓋非排他性的包含，從而使得包括一系列要素的過程、方法、商品或者設備不僅包括那些要素，而且還包括沒有明確列出的其他要素，或者是還包括為這種過程、方法、商品或者設備所固有的要素。在沒有更多限制的情況下，由語句“包括一個……”限定的要素，並不排除在包括所述要素的過程、方法、商品或者設備中還存在另外的相同要素。

本領域技術人員應明白，本發明的實施例可提供為方法、系統或電腦程式產品。因此，本發明可採用完全硬體實施例、完全軟體實施例或結合軟體和硬體方面的實施例的形式。而且，本發明可採用在一個或多個其中包含有電腦可用程式碼的電腦可用儲存媒體(包括但不限於磁碟記憶體、CD-ROM、光學記憶體等)上實施的電腦程式產品的形式。

本發明可以在由電腦執行的電腦可執行指令的一般上

下文中描述，例如程式模組。一般地，程式模組包括執行特定任務或實現特定抽象資料類型的常式、程式、物件、元件、資料結構等等。也可以在分散式運算環境中實踐本發明，在這些分散式運算環境中，由透過通訊網路而被連接的遠端處理設備來執行任務。在分散式運算環境中，程式模組可以位於包括儲存設備在內的本地和遠端電腦儲存媒體中。

本說明書中的各個實施例均採用漸進的方式描述，各個實施例之間相同相似的部分互相參見即可，每個實施例重點說明的都是與其他實施例的不同之處。尤其，對於系統實施例而言，由於其基本相似於方法實施例，所以描述的比較簡單，相關之處參見方法實施例的部分說明即可。

以上所述僅為本發明的實施例而已，並不用於限制本發明。對於本領域技術人員來說，本發明可以有各種更改和變化。凡在本發明的精神和原理之內所作的任何修改、等同替換、改進等，均應包含在本發明的申請專利範圍的範疇之內。

【符號說明】

S102：步驟

S104：步驟

S106：步驟

S108：步驟

S110：步驟

S202：步驟

S204：步驟

S206：步驟

S402：步驟

S404：步驟

S406：步驟

S502：步驟

S504：步驟

S602：步驟

S604：步驟

S606：步驟

700：異常群體識別裝置

701：獲取模組

702：確定模組

703：挖掘模組

704：構建模組

705：聚類別模組

801：處理器

802：記憶體

803：電源

804：有線或無線網路介面

805：輸入輸出介面

806：鍵盤

【發明申請專利範圍】

【第1項】

一種異常群體識別方法，其特徵在於，包括：

獲取多個待分析用戶中的各該待分析用戶的特徵值；

確定各該待分析用戶的特徵值中的高頻特徵值和低頻特徵值；

根據各該待分析用戶的高頻特徵值和預設的頻繁項集挖掘策略挖掘最大頻繁項集，獲取該最大頻繁項集中的低頻最大頻繁特徵值；

根據各該待分析用戶的特徵值中的該低頻最大頻繁特徵值和該低頻特徵值構建目標二部圖，並定義該目標二部圖中的邊的權重；以及

根據該目標二部圖中的邊的權重，以及透過對該目標二部圖進行圖聚類所得到的該多個待分析用戶的聚類結果，確定該待分析用戶中的異常群體。

【第2項】

根據請求項1所述的異常群體識別方法，其中，該獲取多個待分析用戶中的各該待分析用戶的特徵值包括：

獲取該多個待分析用戶的原始個人資料；以及

對該多個待分析用戶的原始個人資料進行離散化，以得到各該待分析用戶的特徵值。

【第3項】

根據請求項1所述的異常群體識別方法，其中，該確定各該待分析用戶的特徵值中的高頻特徵值和低頻特徵值

包括：

根據各該待分析用戶的特徵值構建第一二部圖，其中，該第一二部圖包括與各該待分析用戶對應的節點、與各該特徵值對應的節點、以及各該待分析用戶對應的節點與其特徵值對應的節點之間的邊；

在該第一二部圖中獲取各該特徵值對應的節點的度，並根據各該特徵值對應的節點的度在該特徵值中確定高頻特徵值和低頻特徵值；以及

根據該高頻特徵值和該低頻特徵值確定各該待分析用戶的特徵值中的高頻特徵值和低頻特徵值。

【第4項】

根據請求項1所述的異常群體識別方法，其中，該根據各該待分析用戶的高頻特徵值和預設的頻繁項集挖掘策略挖掘最大頻繁項集，獲取該最大頻繁項集中的低頻最大頻繁特徵值包括：

根據各該待分析用戶的高頻特徵值並結合FP-Growth方法，挖掘支援度滿足預設支援度的頻繁多項集，並在該頻繁多項集中確定最大頻繁項集；

將各該待分析用戶的特徵值與該最大頻繁項集中的最大頻繁特徵值進行匹配，以得到各該待分析用戶的最大頻繁特徵值；以及

在該待分析用戶的最大頻繁特徵值中確定低頻最大頻繁特徵值。

【第5項】

根據請求項4所述的異常群體識別方法，其中，該在該待分析用戶的最大頻繁特徵值中確定低頻最大頻繁特徵值包括：

根據各該待分析用戶的最大頻繁特徵值構建第二二部圖，其中，該第二二部圖包括與各該待分析用戶對應的節點、與各該最大頻繁特徵值對應的節點、以及各該待分析用戶對應的節點與其最大頻繁特徵值對應的節點之間的邊；以及

在該第二二部圖中獲取各該最大頻繁特徵值對應的節點的度，並根據各該最大頻繁特徵值對應的節點的度在該最大頻繁特徵值中確定低頻最大頻繁特徵值。

【第6項】

根據請求項1所述的異常群體識別方法，其中，該根據該目標二部圖中的邊的權重，以及透過對該目標二部圖進行圖聚類所得到的該多個待分析用戶的聚類結果，確定該待分析用戶中的異常群體包括：

在該目標二部圖中刪除權重小於第一預設權重的邊，以得到待聚類二部圖，並對該待聚類二部圖採用聯通演算法得到至少一個最大連通子圖，以及將每個該最大連通子圖中的節點對應的待分析用戶確定為一個該異常群體；或者

在該目標二部圖中刪除權重小於第一預設權重的邊，以得到待聚類二部圖，並透過社區發現演算法對該待聚類二部圖中的節點進行劃分，以得到多個節點集合，以及將

每個該節點集合中的節點對應的待分析用戶確定為一個該異常群體。

【第7項】

根據請求項1所述的異常群體識別方法，其中，該根據該目標二部圖中的邊的權重，以及透過對該目標二部圖進行圖聚類所得到的該多個待分析用戶的聚類結果，確定該待分析用戶中的異常群體包括：

根據該目標二部圖中的邊的權重計算任意兩個該待分析用戶之間的權重；

將各該待分析用戶轉化為節點，並在任意兩個節點之間設置邊，並將任意兩個節點的邊的權重設定為對應的任意兩個該待分析用戶之間的權重，以構建目標聚類圖；以及

透過對該目標聚類圖進行圖聚類所得到的該多個待分析用戶的聚類結果，確定該待分析用戶中的異常群體。

【第8項】

根據請求項7所述的異常群體識別方法，其中，該透過對該目標聚類圖進行圖聚類所得到的該多個待分析用戶的聚類結果，確定該待分析用戶中的異常群體包括：

在該目標聚類圖中刪除權重小於第二預設權重的邊，以得到待聚類圖，並對該待聚類圖採用聯通演算法得到至少一個最大連通子圖，以及將每個該最大連通子圖中的節點對應的待分析用戶分別確定為一個該異常群體；或者

在該目標聚類圖中刪除權重小於第二預設權重的邊，

以得到待聚類圖，並透過社區發現演算法對該待聚類圖進行劃分，以得到多個節點集合，以及將每個該節點集合對應的待分析用戶分別確定為一個該異常群體。

【第9項】

一種異常群體識別裝置，其特徵在於，包括：

獲取模組，用於獲取多個待分析用戶中的各該待分析用戶的特徵值；

確定模組，用於確定各該待分析用戶的特徵值中的高頻特徵值和低頻特徵值；

挖掘模組，用於根據各該待分析用戶的高頻特徵值和預設的頻繁項集挖掘策略挖掘最大頻繁項集，獲取該最大頻繁項集中的低頻最大頻繁特徵值；

構建模組，用於根據各該待分析用戶的特徵值中的該低頻最大頻繁特徵值和該低頻特徵值構建目標二部圖，並定義該目標二部圖中的邊的權重；以及

聚類別模組，用於根據該目標二部圖中的邊的權重，以及透過對該目標二部圖進行圖聚類所得到的該多個待分析用戶的聚類結果，確定該待分析用戶中的異常群體。

【第10項】

一種異常群體識別設備，其特徵在於，包括：

處理器；以及

被安排成儲存電腦可執行指令的記憶體，該電腦可執行指令在被執行時使該處理器：

獲取多個待分析用戶中的各該待分析用戶的特徵值；

確定各該待分析用戶的特徵值中的高頻特徵值和低頻特徵值；

根據各該待分析用戶的高頻特徵值和預設的頻繁項集挖掘策略挖掘最大頻繁項集，獲取該最大頻繁項集中的低頻最大頻繁特徵值；

根據各該待分析用戶的特徵值中的該低頻最大頻繁特徵值和該低頻特徵值構建目標二部圖，並定義該目標二部圖中的邊的權重；以及

根據該目標二部圖中的邊的權重，以及透過對該目標二部圖進行圖聚類所得到的該多個待分析用戶的聚類結果，確定該待分析用戶中的異常群體。

【第11項】

一種儲存媒體，用於儲存電腦可執行指令，其特徵在於，該電腦可執行指令在被執行時實現以下流程：

獲取多個待分析用戶中的各該待分析用戶的特徵值；

確定各該待分析用戶的特徵值中的高頻特徵值和低頻特徵值；

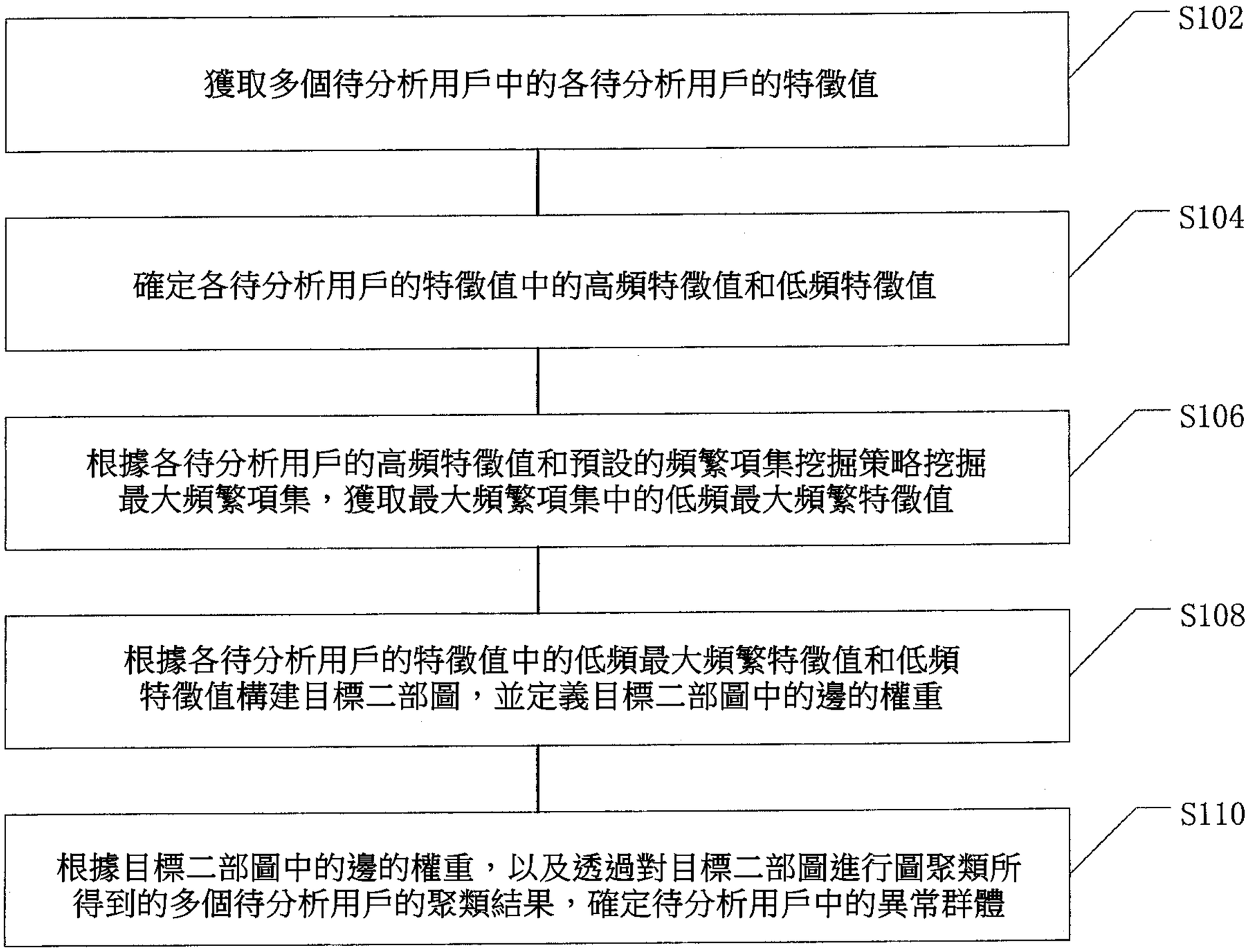
根據各該待分析用戶的高頻特徵值和預設的頻繁項集挖掘策略挖掘最大頻繁項集，獲取該最大頻繁項集中的低頻最大頻繁特徵值；

根據各該待分析用戶的特徵值中的該低頻最大頻繁特徵值和該低頻特徵值構建目標二部圖，並定義該目標二部圖中的邊的權重；以及

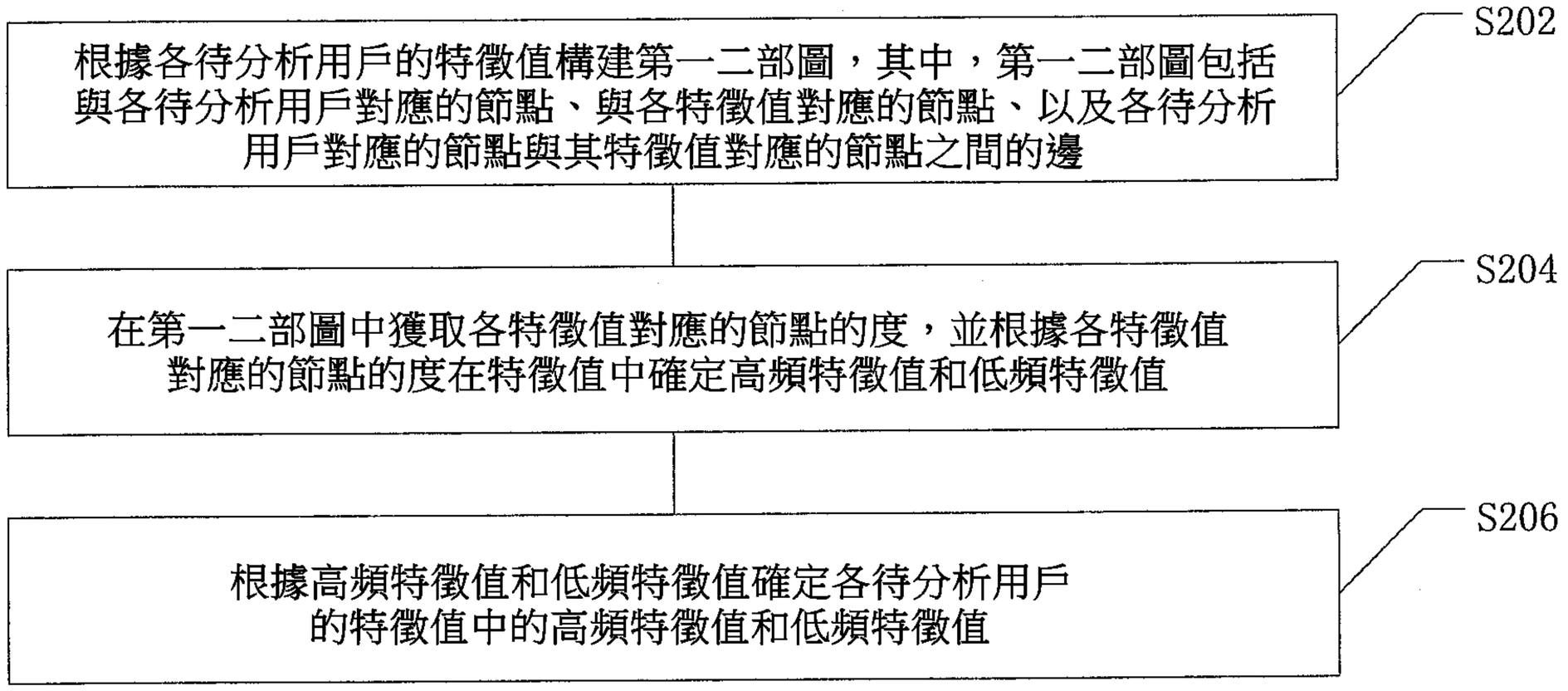
根據該目標二部圖中的邊的權重，以及透過對該目標

二部圖進行圖聚類所得到的該多個待分析用戶的聚類結果，確定該待分析用戶中的異常群體。

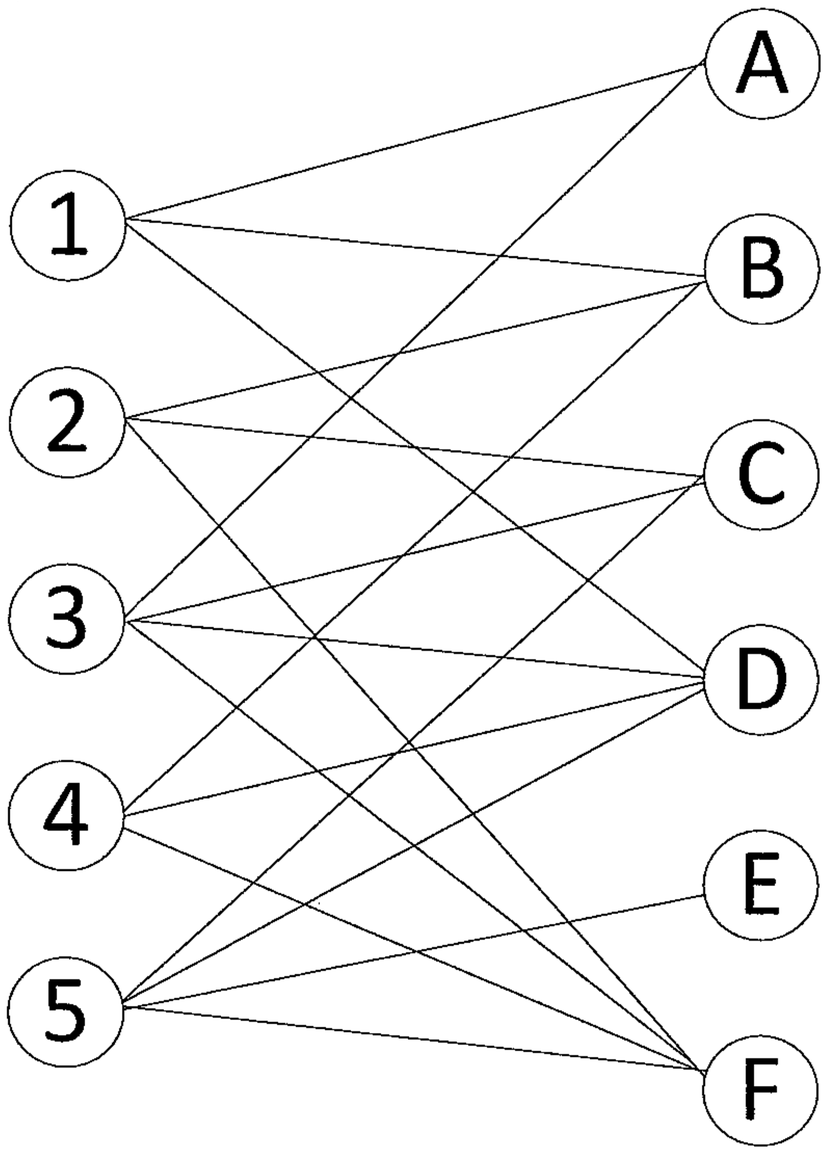
【發明圖式】


S102
獲取多個待分析用戶中的各待分析用戶的特徵值
S104
確定各待分析用戶的特徵值中的高頻特徵值和低頻特徵值
S106
根據各待分析用戶的高頻特徵值和預設的頻繁項集挖掘策略挖掘最大頻繁項集，獲取最大頻繁項集中的低頻最大頻繁特徵值
S108
根據各待分析用戶的特徵值中的低頻最大頻繁特徵值和低頻特徵值構建目標二部圖，並定義目標二部圖中的邊的權重
S110
根據目標二部圖中的邊的權重，以及透過對目標二部圖進行圖聚類所得到的多個待分析用戶的聚類結果，確定待分析用戶中的異常群體

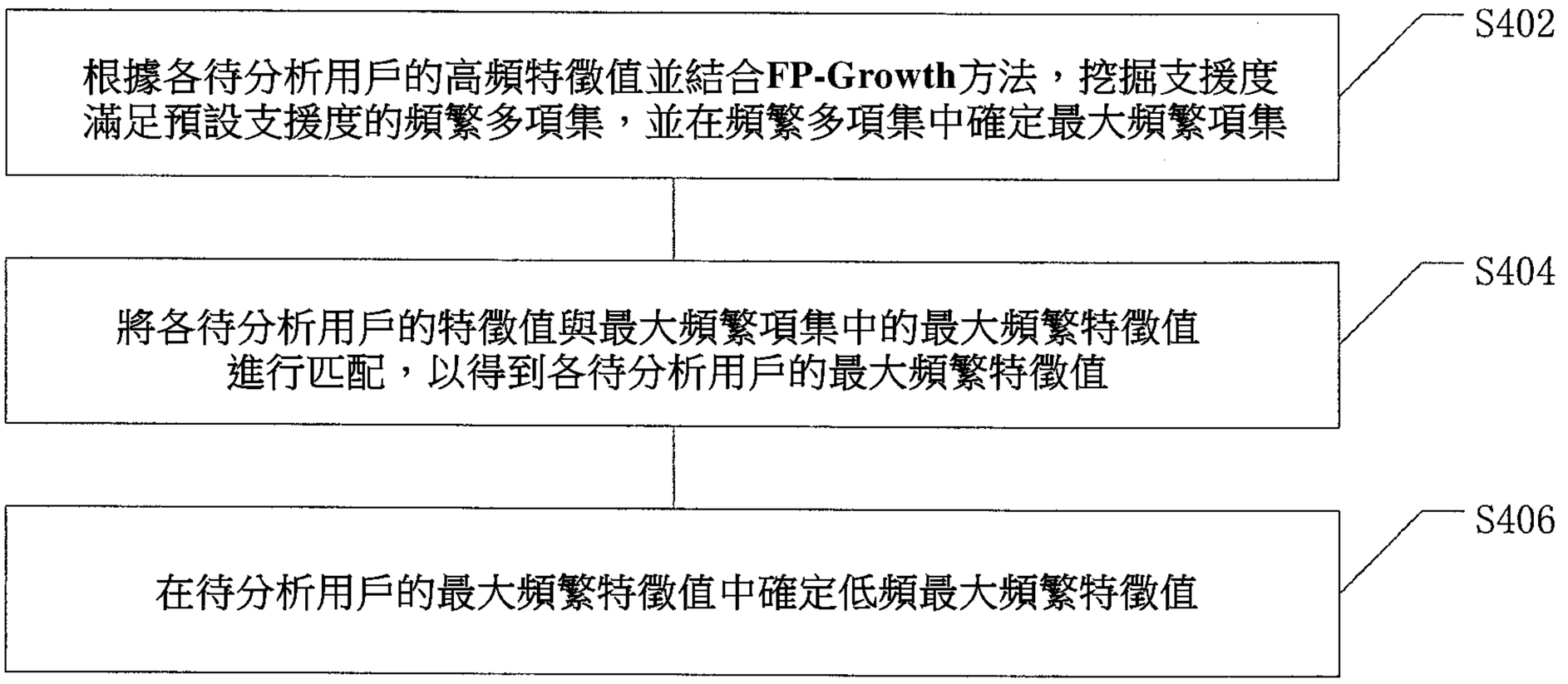
【圖 1】


S202
根據各待分析用戶的特徵值構建第一二部圖，其中，第一二部圖包括與各待分析用戶對應的節點、與各特徵值對應的節點、以及各待分析用戶對應的節點與其特徵值對應的節點之間的邊
S204
在第一二部圖中獲取各特徵值對應的節點的度，並根據各特徵值對應的節點的度在特徵值中確定高頻特徵值和低頻特徵值
S206
根據高頻特徵值和低頻特徵值確定各待分析用戶的特徵值中的高頻特徵值和低頻特徵值

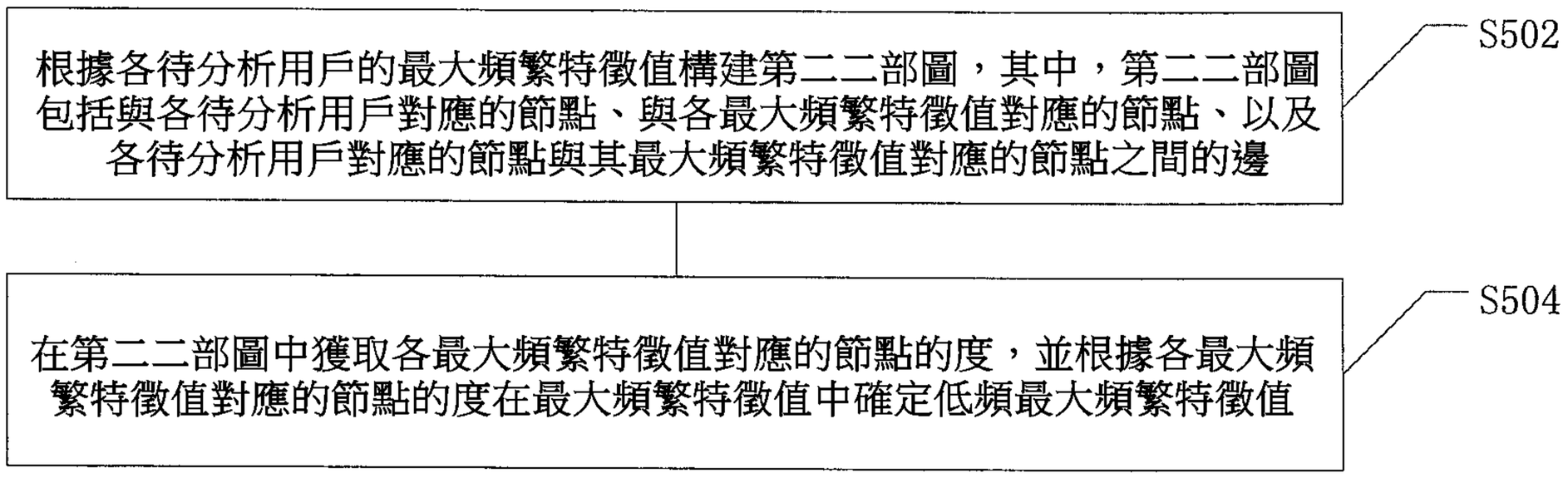
【圖 2】


1
2
3
4
5
A
B
C
D
E
F

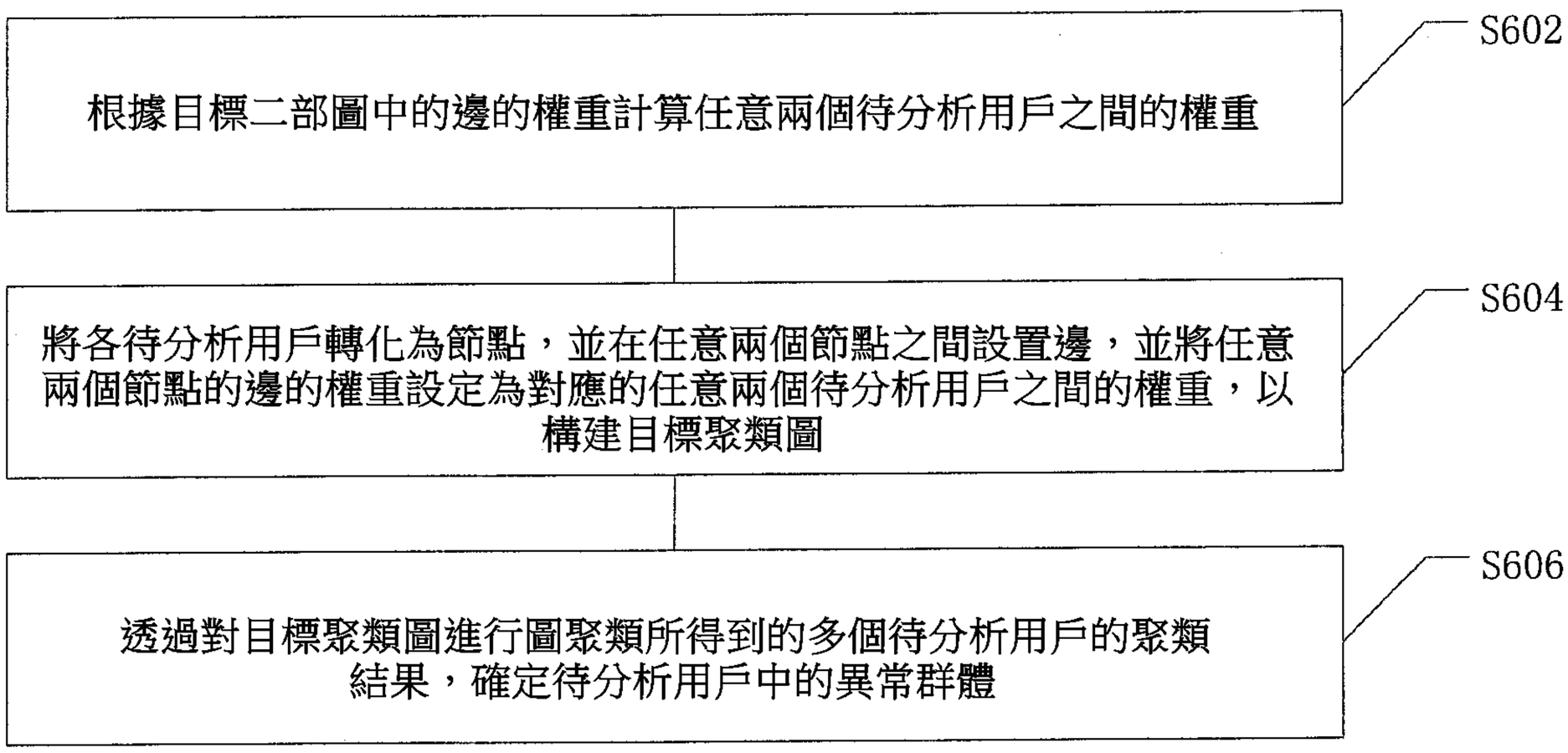
【圖 3】


根據各待分析用戶的高頻特徵值並結合FP-Growth方法，挖掘支援度滿足預設支援度的頻繁多項集，並在頻繁多項集中確定最大頻繁項集
S402
將各待分析用戶的特徵值與最大頻繁項集中的最大頻繁特徵值進行匹配，以得到各待分析用戶的最大頻繁特徵值
S404
在待分析用戶的最大頻繁特徵值中確定低頻最大頻繁特徵值
S406

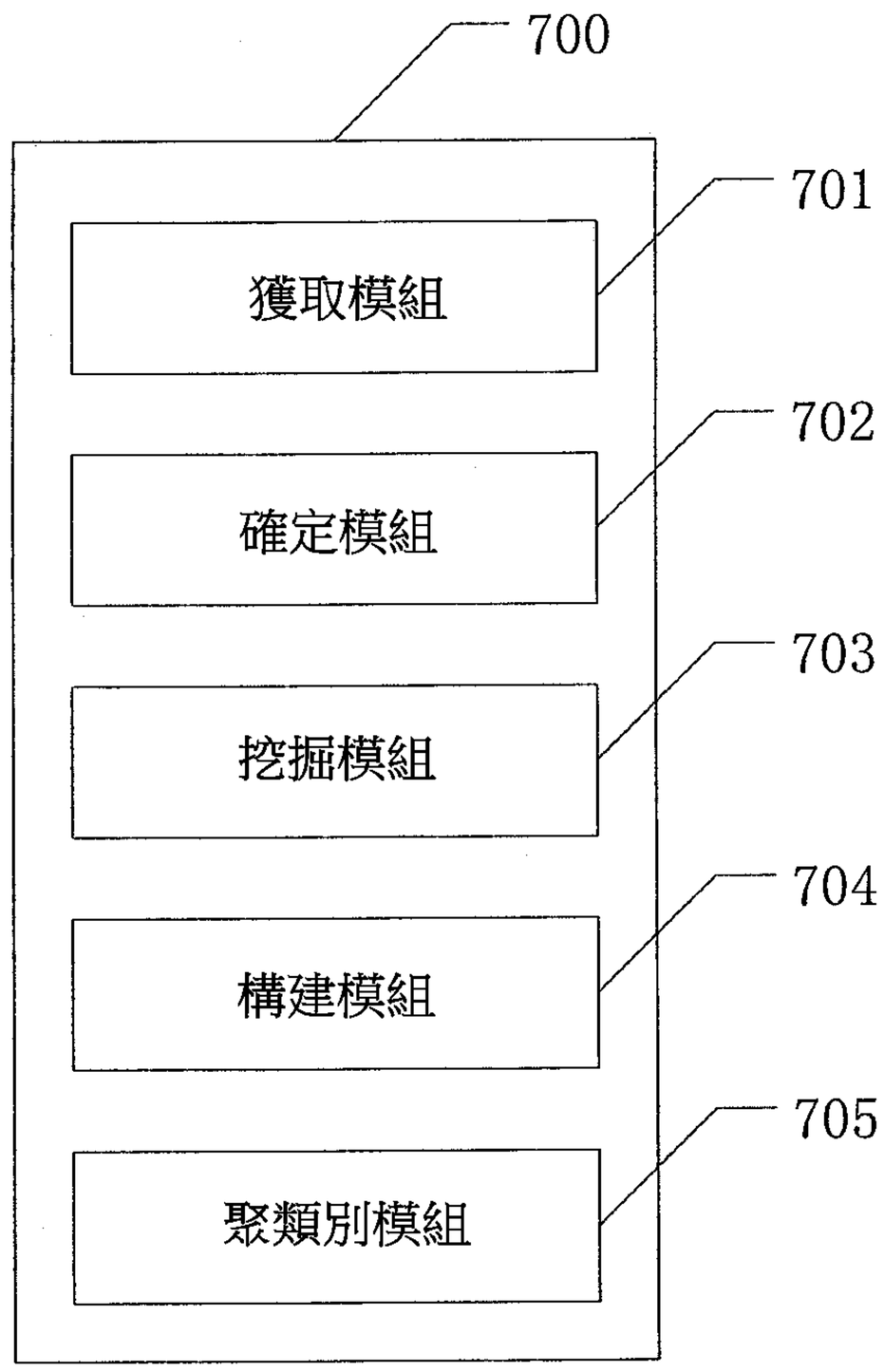
【圖 4】


根據各待分析用戶的最大頻繁特徵值構建第二二部圖，其中，第二二部圖包括與各待分析用戶對應的節點、與各最大頻繁特徵值對應的節點、以及各待分析用戶對應的節點與其最大頻繁特徵值對應的節點之間的邊
S502
在第二二部圖中獲取各最大頻繁特徵值對應的節點的度，並根據各最大頻繁特徵值對應的節點的度在最大頻繁特徵值中確定低頻最大頻繁特徵值
S504

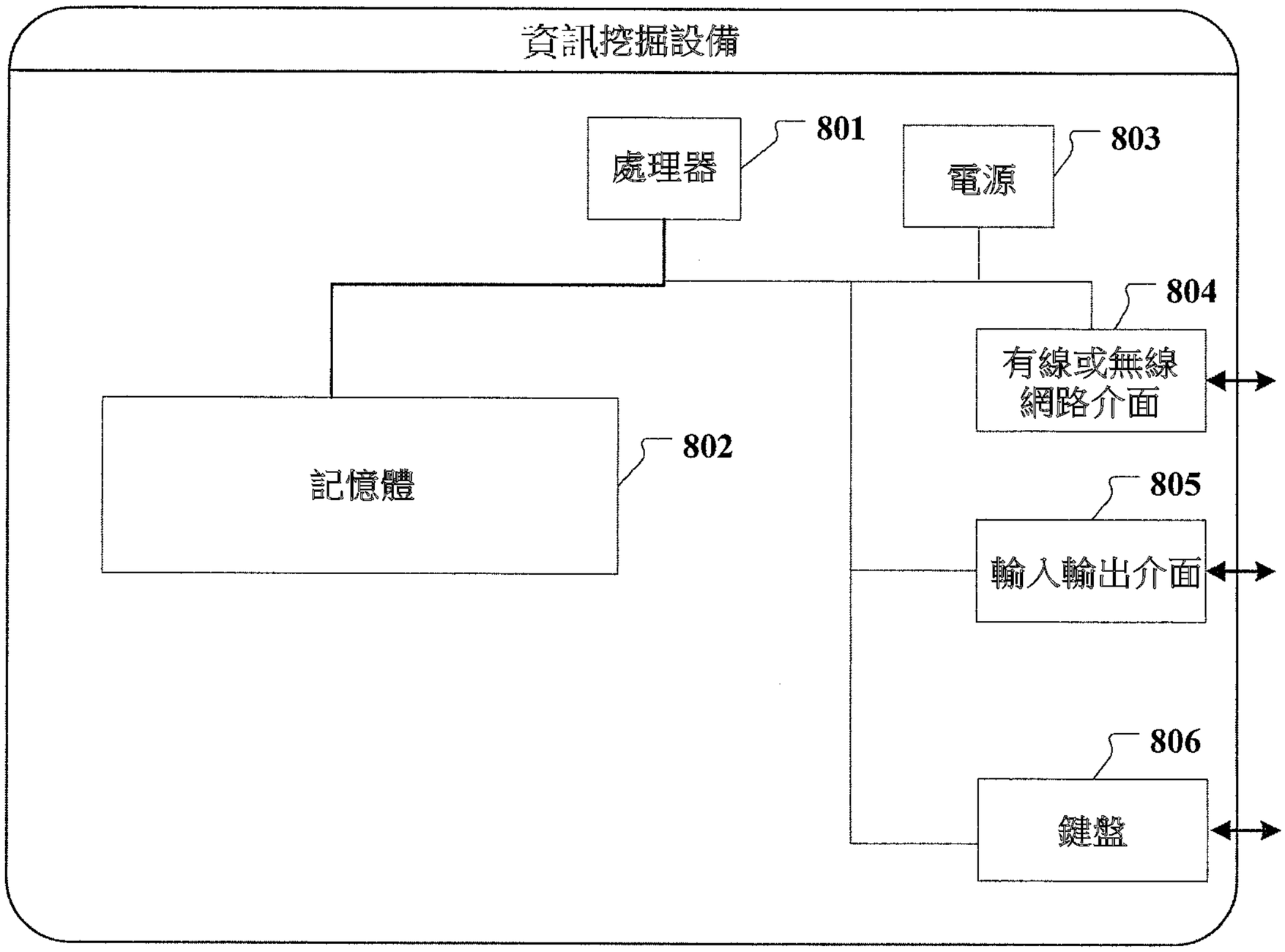
【圖 5】


根據目標二部圖中的邊的權重計算任意兩個待分析用戶之間的權重
S602
將各待分析用戶轉化為節點，並在任意兩個節點之間設置邊，並將任意兩個節點的邊的權重設定為對應的任意兩個待分析用戶之間的權重，以構建目標聚類圖
S604
透過對目標聚類圖進行圖聚類所得到的多個待分析用戶的聚類結果，確定待分析用戶中的異常群體
S606

【圖 6】


700
獲取模組
701
確定模組
702
挖掘模組
703
構建模組
704
聚類別模組
705

【圖 7】



【圖 8】