DataMining Project3

P76121186 楊祐昇

1. Find a way
2. Algorithm description

首先先定義兩個class分別為Graph(儲存一個圖所有的點)、Node(儲存一個點該有的內容)，再根據3個演算法所需的變數與功能去擴充

一張含有 文字, 字型, 螢幕擷取畫面, 數字 的圖片

自動產生的描述

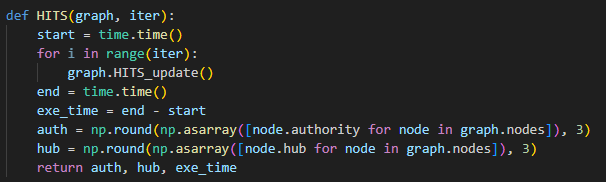
一個圖會有一個List裡面放所有的點(self.nodes)、similarity matrix，多設置一個new\_sim\_matrix是為了避免再做similarity matrix 更新時儲存到該輪已更新的內容，所以將更新內容先存至new\_sim\_matrix，decay\_factor則為更新similarity matrix需要用到的參數

一張含有 文字, 螢幕擷取畫面, 字型 的圖片

自動產生的描述

name為結點的名稱，children、parents分別儲存此點的下一層及上一層節點，另外每個點都會有authority、hub、pagerank三個值，另外具有new的變數同樣也是為了避免計算新的authority、hub、pagerank時用到已被更新的值，在每輪更新時先存在這些變數中

* HITS



根據iter參數決定要執行幾次update，exe\_time為update iter 次所需的時間，將authority、hub結果四捨五入至小數點第三位並且回傳

一張含有 文字, 螢幕擷取畫面, 字型 的圖片

自動產生的描述一張含有 文字, 螢幕擷取畫面, 字型 的圖片

自動產生的描述

* HITS每次update(左圖)

1. 針對每個點計算其新authority值並存在new\_authority中，計算方法(右圖updatae\_auth)為將所有射入點的hub值相加即為新authority值
2. 針對每個點計算其新hub值，計算方法(右圖update\_hub)為將所有射出點上一輪的authority值相加即為新hub值
3. 將new\_authority存到authority中做為下一輪的上一輪值使用
4. 最後將所有點的authority、hub正規化，結束此次update

一張含有 文字, 螢幕擷取畫面, 字型 的圖片

自動產生的描述

* HITS的正規化

1. 計算所有點新的authority、hub值的和
2. 將每個點的authority、hub都除step1計算出來的結果即完成normalize

* PageRank

一張含有 文字, 螢幕擷取畫面, 字型 的圖片

自動產生的描述

1. 先將所有nodes根據其name排序
2. 因為pagerank初始值為1/N，但創建graph每個node的pagerank預設為1，所以必須將每個點的pagerank / 總點數(使用init\_pagerank()初始化)
3. 根據iter決定pagerank要更新幾次
4. 最後將所有點之pagerank四捨五入至小數點第三位連同執行時間回傳

一張含有 文字, 字型, 螢幕擷取畫面, 圖形 的圖片

自動產生的描述一張含有 文字, 字型, 螢幕擷取畫面 的圖片

自動產生的描述

一張含有 文字, 字型, 螢幕擷取畫面 的圖片

自動產生的描述

* PageRank update

1. 針對所有點去更新其pagerank值(如何更新如下圖)
2. 將所有點之pagerank值做正規化結束一輪

一張含有 文字, 螢幕擷取畫面, 字型 的圖片

自動產生的描述

一張含有 字型, 文字, 白色, 行 的圖片

自動產生的描述

* 如何更新pagerank

1. 將所有入射點的pagerank加總(每個入射邊的pagerank值為入射點的pagerank值 / 入射點的射出數量)
2. 最後將加總的值根據上圖公式算出新的pagerank值並儲存在new\_pagerank，避免之後的點在更新時儲存到該輪已更新的值

一張含有 文字, 螢幕擷取畫面, 字型 的圖片

自動產生的描述

* Pagerank 正規化

1. 計算所有點pagerank加總，並將new\_pagerank值賦予pagerank
2. 將所有點除Step1加總結果即完成normalization

* SimRank

一張含有 文字, 螢幕擷取畫面, 字型, 多媒體軟體 的圖片

自動產生的描述

1. 初始化SimRank matrix
2. 根據iter決定要更新 SimRank幾次
3. 最後將SimRank matrix四捨五入至小數點第三位連同執行時間回傳

一張含有 文字, 螢幕擷取畫面, 字型 的圖片

自動產生的描述

* 初始化SimRank matrix

先將各點排序，建立matrix，對角線為1其餘為0

一張含有 文字, 字型, 螢幕擷取畫面 的圖片

自動產生的描述

一張含有 文字, 螢幕擷取畫面, 軟體 的圖片

自動產生的描述

* Similarity值更新

針對每SimRank每格計算兩點的Similarity值

* 如何計算兩點similarity值

一張含有 字型, 文字, 筆跡, 書法 的圖片

自動產生的描述

* 1. 根據兩點的parents求出所有組合上一層similarity值加總

2. 最後乘上(decay factor) / 兩點parents數相乘即為兩點新的similarity值並存在new\_sim\_matrix中避免其他格更新時存到已更新的值

1. Result analysis and discussion
2. Effectiveness analysis