國立中興大學論文寫作範本

使用說明

1. 此檔案僅供研究生套用格式，論文主體架構請依「[國立中興大學學位論文格式規範](http://www.nchu.edu.tw/~regist/download/download/thesisModify981126.doc)」規定撰寫。
2. 論文完成後，請**務必刪除**「本頁使用說明」、出現在各頁左上角的「文字方塊」及「註解」(點選[校閱]索引標籤 後，點選[顯示註解]查看)。

國立中興大學資訊科學與工程學系

**書名頁**

碩士學位論文

基於k鄰近填補法在不完整資料集中

找尋近似天際線

Finding approximate skyline set via

k-NN based imputation under incomplete data set

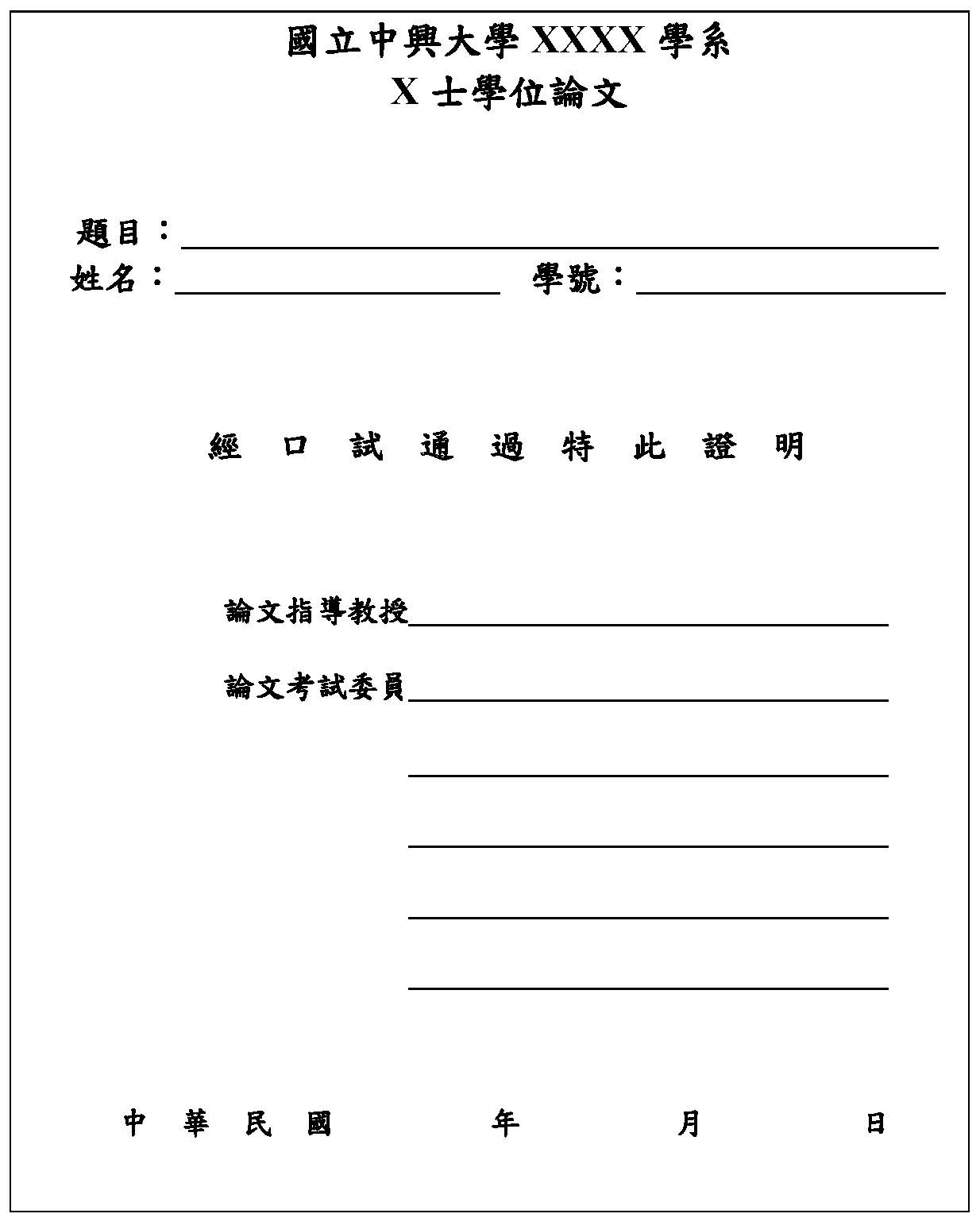
指導教授︰賈坤芳　 Kuen-Fang Jea

研 究 生︰凌政楠　 Cheng-Nan Ling

中華民國一百零X年XX月

說明：由各系（所）提供，經口試委員、指導教授簽名，以證明論文口試合格。樣式請見如下：

**審核頁**



**誌謝辭**

說明：表達對師長、受訪者、同學、家人等感謝之意，以一頁為原則，最多不超過兩頁。(非必備，由各系所自行決定)

# 摘要

說明：以一頁為原則，最多不超過兩頁，簡要說明研究動機、研究方法與設計、資料收集與分析、研究結果及討論建議等。

關鍵字：建議從論文題目或摘要中挑選1-2個關鍵字。

中文摘要

# Abstract

說明：精簡原則與中文摘要相同，必須注意專有名詞的翻譯是否正確。

英文摘要

Keywords:1-2 keywords

# 目次

[摘要 i](#_Toc532476759)

[Abstract ii](#_Toc532476760)

[目次 iii](#_Toc532476761)

[表目次 iv](#_Toc532476762)

[圖目次 v](#_Toc532476763)

[第一章 1](#_Toc532476764)

[參考書目 3](#_Toc532476765)

[一、中文部分 3](#_Toc532476766)

[二、西文部分 3](#_Toc532476767)

[附錄 4](#_Toc532476768)

說明：按「國立中興大學學位論文格式規範」之「論文編印項目次序」所列各項順序，依次註明章節編號、內容項目名稱及所在頁次，右邊對齊。

# 表目次

說明：註明圖表的編號、名稱及所在頁次，右邊對齊。表在前，圖在後。如圖或表總數超過10個以上，則分列表目次及圖目次。

# 圖目次

說明：原則與表目次相同。

# 第一章

正文

## 1.1 簡介

現實生活中，在蒐集資料的過程當中難免會面臨到資料不齊全的狀況，可能在蒐集過程中不慎讓內容流失，亦有可能是蒐集到的資料值很難完備，而此一類資料集在資料科學領域中被稱之為「不完整資料集」，即資料當中不全具有值。

在多維度資料集當中，資料難免會因為諸多因素造成資料集內有缺失的情形從而影響結果，而針對缺失值填補的演算法其中又以k鄰近填補法對不同缺失狀況表現較好，然而，過往的k鄰近填補法都是以一個自定義常數k作為對所有缺失值找到該維度鄰近點的方法，但卻少有研究針對不同缺失情形下做個別處理，一般而言，在各個維度當中無論缺失值的分布狀況較如何，一旦採取k鄰近填補法機制填補法，在所有的維度中，缺失個數、某一缺失值與該維度所佔有比例等等，均予以同一機制搜尋k個鄰近點後，並在此方法下所找到的鄰近點也予以相同權重，將值填補回該缺失值欄位。

根據上述之情形，在以往經過傳統k鄰近填補法填補後的效果可能不會那麼好，因此，本研究想觀察在對於缺失值分布不同的情況以及不同缺失情形下，基於原k鄰近填補方法上做改善，藉以達到可以修正考慮某些缺失分布情形下，填補後的值可以更加合理且完善，並提出一個改進的方法。

# 第二章 相關研究

本論文所牽涉到的相關議題有4個面向 :  
天際線問題(skyline problem)、缺失值處理(missing value handling)、填補法 (imputation)、k鄰近填補法(k-nearest neighbor imputation)

## 天際線問題 Skyline set problem

### 天際線問題概述與完整資料集對天際線問題之影響

天際線問題主要目的是要找出所有不會被任何一點支配的資料點集合，其決定被支配與否的關鍵是在比較點p的所有維度下的值均不小於或等於點q，稱之為點p支配點q，但由於天際線問題必須比較所有的維度方能決定支配性與否，故一個完整資料集對天際線問題具有舉足輕重的影響力，在缺失值的資料集下欲比較所有維度是一項重大的困難所在。

## 2.2 缺失值處理技術 Missing value handling

### 2.2.1 資料缺失類型

首先處理缺失值技術前，先簡單分類缺失值的種類

隨機缺失值Missing at Random (MAR):

在分析當中控制了其他變數，缺失的機率與該欄位無關，則此一類型缺失成為隨機缺失。

(舉例說明)

Missing Completely at Random (MCAR):

假設一變數有缺失數據，若該缺失數據的機率與該維度本身的值或該數據中任何其他店數的值都沒有相關，則可稱此一缺失類型為完全隨機缺失

(舉例說明)

Missing not at Random (MNAR):

均非屬於上述兩者，表示缺失資料值與該維度具有一定程度的關係或傾向，屬於此一類型缺失資料會表現某一種資料特性，故又被稱作不可忽略缺失類型。

(舉例說明)

### 2.2.2 資料集缺失值的處理技術

根據以往經驗當中，處理缺失值資料不外乎以下幾種 :

1. 丟棄 (dropout)

1-1. 刪除資料列

如果一個資料列裡面的欄位當中有觀察到具有缺失值，則將整筆資料列直接刪除。

1-2. 刪除該特徵欄位

若整體資料集的某一特徵欄位缺失筆數的數量太多，甚至遠多於其他特徵欄位缺失狀況時，則可選擇直接放棄該特徵欄位。

遇缺失值時無論就上述採用丟棄或刪除法，雖然可能不需要面對因具有缺失值所造成的資料不完整性，然而另一方面無論是丟棄資料列或是丟棄整個特徵欄位，仍然需要面臨因丟棄所產生的額外問題 : 喪失資料特徵性質。

1. 填補法

用最簡單的統計方法來計算需要放入的插補值，常見的補值方法有給予單一固定值、眾數、平均值、中位數......等等

為避免遺失資料點在整體資料集所表現的特性與影響統計數字，無論是丟棄該資料列或是刪除整個特徵欄位，整體而言填補法還是會有比較好的影響，本研究亦是屬於填補法的技術處理缺失值。

## 2.3 填補法 Imputation

眾多填補法當中又可以分為單一填補法(single imputation)與多重填補法(multiple imputation)，兩者

### 2.3.1 單一填補法與多重填補法

多重填補法

相較於單一填補法，多重填補法更著重於分析與解決問題上，多重填補法仰賴於資料集上模擬模型，在遇到缺失資料時會根據模型給予一群可能為解的解集合，並在陸續填補過程中，調整資料集的分布、變異數以及信賴區間等，因此在填補過程中會有大量的計算需求。

單一填補法

單一填補法主要目的是填補某一特定空缺上被遺失的值，即使被計算出可能的解不只一種也會隨機挑選其中一個值作為最後填補的結果，當缺失資料在整你資料集佔有很少的數量時，單一填補法則不失為一種簡易卻又實用的方式，又因為這種填補法計算量小，最常被用來當作填補缺失值的首選方法，另外，單一填補法在面臨一定程度的缺失情形時，可能會面臨到一個嚴重的問題，一旦被填補進缺失欄位後，便會被當作真正的資料值而無法在分辨出原始資料集的值與被填補值的真偽，甚至可能造成日後誤導分析結果，而k鄰近填補法則屬於此一類型填補法。

## 2.4 k鄰近填補法 k-Nearest Neighbor imputation

### 2.4.1 探討k鄰近填補法的優、缺點

k-nearest neighbor imputation (往後簡稱kNN 或 k鄰近填補法)至今仍是一個很實用且廣為流傳的填補缺失值方法，k鄰近填補法的核心概念即在多維度空間資料集中某參考點p，藉著找尋其他k個與p點相鄰最接近的點在p點缺失的維度上作為填補點p的參考值，k鄰近填補法可以被應用於的資料種類不僅僅是連續型資料(continuous data)上，也可以被應用於離散型資料(discrete data)、有序型資料(ordinal)甚至是分類型資料(categorical data)，幾乎均適用在各式各樣的資料種類。

k鄰近填補法為上述所有補值方法當中，最為常見也是普遍被認為效果比較好的補值法，以下概述其優、缺點 :

kNN優點

一般而言，kNN會比單一填補值，如:平均數、中位數、極值、或是眾數等填補法還要來得準確許多，原因是該方法會同時參照其他該缺失值相鄰點去預測其應該原有的合理值，尤其當最終必須找出skyline set的點時，單一值填補法對於尋找skyline point set不會有更好的幫助，反而會因此增加許多不必要的計算在比較相同維度上具有相同值，因缺失而被填補回去的相同值。

類似這樣的問題在無論考慮單一維度或整體資料集缺失值密度，當missing rate 愈趨增加時，其餘具有完整資料值的點也會越來越少。

總然如此，kNN 也存在者某些潛在缺點

身為非監督式填補法會參考其餘非缺失值，換句話說，此類型填補法並不會對資料內容做任何過濾或預處理，對於既有資料敏感度距甚，很容易受限於輸入資料集當中非缺失值的內容是否為noise value導致結果很有可能無法精準地將資料集內的原缺失部分填補回去。

另外也不需考量的點是，現今實際層面常常處理資料量大而且又高維度的巨量資料，k鄰近填補法必須先行儲存整個欲參考的鄰近點於記憶體中，並且逐一計算求出各點兩兩之距離，可想而知其計算量所費不貲。

對某一個點p來說，在找尋其相鄰近點集合過程裡，其所有相鄰近點之中從第一個至第k個相鄰近點都具有相同的權重值，但是在計算點p與資料集當中其他點q的與點p距離同時，並且對所有距離值先行排序，如此距離值的排序行為已經代表著點p與點q的距離上的意義，我們應該利用這樣的排序關係設計出一個合理的權重比值，以表示被參考的其他點q在填補的演算過程中應該被參考的價值有多少分量，最常見也最簡易的權重表示方式即為兩點距離的倒數。

# 第三章 問題與方法

## 3.1 符號定義

|  |  |
| --- | --- |
| notation | description |
| n | number of data instances within a input data set |
| m | dimensionality of input data set |
| Incomplete data set | an incomplete data set with certain missing values as input data set,  with size |
|  | data value of the data sample,  at index row and column |
|  | data instance at row, composed by |
| Imputed data set | a completed data set with missing value all imputed, with size |
| , k | specified constant k |
|  | a weight matrix records pairwise any and , i |
|  | type of weighting, determine mechanism of the way assigning weight in |
| = | weighting value with respect to between and |
|  | a distance matrix is a symmetric matrix, which records the distance between pairwise two data sample |
|  | distance between and , denoted as , |
| NN list | record a sorted nearest neighbor list between all pairs of and |
|  | a sorted list at index i in NN list  keep all neighbors of , which |
|  | nearest neighbor of  element at index j in sorted list |
|  | value been imputed after imputation at index (i, j) only if is missing or value NaN in advanced, otherwise = |
| mask | a boolean data type array with size of k  to label the candidate at column j has non-NaN value |

## 3.2 研究動機

搜尋skyline set時會需要去比較所有特徵欄位的值，也就是說每一筆資料列的每一個特徵欄位都必須有值存在，因此，具有完整的資料集是 skyline problem中必要的先決條件。但在實際情況下，期待蒐集每一個特徵欄位都不能有缺失值似乎是不切實際地，現實生活當中具有太多不可抗拒因素使我們在取得資料集的過程難免會遇到欄位裡的值會有缺失，一般解決缺失值的方法不外乎刪除與丟棄法，但此一類型方式最大的缺點是當missing rate不高的時候，很容易遺失該筆表現在整體資料集所具有的資料特徵，此方法在missing rate低於某一門檻值時，本論文不建議貿然刪除該資料欄位，反之本論文方法為盡可能地保留具有資料特性的機會，~~故此設立一門檻值為80%~~，缺失值則可以0、null或是NaN作為該缺失欄位的表示法。

而在眾多填補法，其中又以k鄰近填補法在大多數情況下填補後的結果普遍表現不錯，原因是k鄰近填補法會參考著其鄰近鄰居的再取算術平均值，使得其所填補後的值相較於其他只以單一數值填補法更具有參考價值。

## 3.3 問題定義

在不完整資料集中，如何改善k鄰近填補法填補缺失值後，使填補後的完整資料集具有最近似skyline set的內容。

## 3.4 問題分析

所有計算距離公式中，歐式距離的計算方法是採資料集中兩兩資料點相對應維度的差平方再取平方根，若是遇到相同維度下至少一個具有NaN值，則在計算歐式距離時並不會採計具有NaN值的數值差平方，此計算方式是最廣為主流的算法，在此論文中並不打算更改計算方式，但由此計算方式可看出淺在問題 :

具有缺失值的兩資料點計算可能會誤導兩資料點間的實際距離

情境一、看似近的兩點實際上距離更遠

情境二、看似遠的兩點實際上距離更近

k鄰近填補法的另一個問題在於，當缺失值越高的時候，當k值很大時意味著鄰近點仍存有non-NaN value的機會並不大，而此時k鄰近點填補法則在找不足k個鄰近點情況下，選擇不從剩下的鄰近點補足，這樣現象尤其當存在nan-NaN值很稀少時更為嚴峻，導致k鄰近填補法會幾乎用同一數值填補回去，如此便會與單純填補單一數值(平均數、眾數、極大值、極小值)無異，填補後找尋天際線時因為該維度幾乎都是同一數值，更容易形成有如該欄位直接被刪除一樣無意義比較的結果。

## 3.5 sk-NN imputation 方法

sk-NN imputation Algorithm(){

|  |
| --- |
| Input : incomplete data set , constant k, weight type  Output : complete data set  step 1. load incomplete data set  step 2. make a copy of , complete data set to be imputed later  step 3. initialize all values of distance matrix into zero  3-1. for each row in distance matrix  3-2. for each column in distance matrix  3-3. = euclidean distance of pairwise data samples    step 4. initialize all values of weight matrix into zero  4-1. for each row in weight matrix  4-2. for each column in weight matrix  4-3. according to weight type , alter the weighting mechanism  =  4-4. return weight matrix  step 5. establish a nearest neighbor list to store all nearest neighbors with respect to certain data sample row  5-1. initialize an empty NN list, with size n  5-2. for each in NN list  5-3. check each in  5-4. retrieve all between any pair of the and from  5-5. sorting in ascending order, keep track of corresponding  5-6. repeat from step 5-6. to step 5-9.  5-7. if 0  5-8. append into  5-9. until size of or all inserted into  5-10. update  5-11. return NN list  step 6. search all missing values among data set, then impute new value back into the missing position  6-1. for each row i in  6-2. for each column j in  6-3. if missing value or value NaN is found  6-4. Impute\_Process(i, j, k)  step 7. return an imputed complete data set |

}

Subroutine, named Impute\_Process

Impute\_Process(i, j, k){

|  |
| --- |
| Input : row index i, column index j, nearest neighbor k  Output : imputed value at index (i, j)  step 1. create a mask array with size k and initialize all False  step 2. find all elements in mapping to has non-NaN value  2-1. for each index from 0 to k  2-2. r =  2-3. if is missing or value NaN  2-4. mask[r] = True  2-5. otherwise  2-6. mask[r] = False    step 3. check Boolean value in mask array  3-1. for each index in mask  3-2. if mask array is not all True  3-3. retrieve all value in  3-4. else  3-5. reset mask array all False  3-6. randomly sample rest non-NaN value at column j  3-7. evaluate mean or weighted mean of retrieved value  3-8. assign imputed value to  step 4. return imputed value |

}

## 3.6 以skyline set作為填補法的表現優劣

# 第四章 實驗與結果分析

## 4.1 實驗環境設定與資料來源

### 4.1.1 實驗環境、資料來源

編寫程式語言 : Python 3.8.2  
撰寫程式環境 : anaconda jupyter Lab, Notebook  
環境引用套件 : pandas, numpy, sklearn, matplotlib  
本實驗資料集 : [dataset source], data\_generator

### 4.1.2 實驗設計

## 4.2 實驗一

### 4.2.1 實驗目的與設計

本實驗目的是在探討參考鄰近點k值越大，對填補後的資料集找skyline set是否會有更好的填補效果。

### 4.2.2 實驗方法

本實驗方法為將同一缺失資料集分別從k=1測試，觀察相同k值分別於不同缺失比例下情況之下，隨著k值增加是否可以得到更佳的準確度。

衡量此實驗效果，本論文採用原完整資料集 complete data set中所得出 skyline set 的點作為最終填補效果的依據。

分析不同鄰近點 k 值對不同缺失值情況下對k-NN imputation的影響

### 4.2.3 實驗結果、分析與實驗結論

根據圖4.1實驗一顯示結果，隨著缺失比率在資料集當中增加，k鄰近填補法的準確率並沒有因為找尋更多的鄰近點數量改善填補效果。

觀察圖4.2可知，k鄰近填補法只著重在將鄰近參考點的數量一味地增加，但因為缺失比例也逐漸增加，再配合k鄰近填補法中對可供參考點不足從缺的機制，致使即使計算鄰近點值的平均也會逐漸失效，這同時也意味著可供參考點數量以及參考值之可靠性嚴重不足。



圖4.1 hit ratio versus miss rate when k = 1



圖4.2 hit ratio versus miss rate when k = 4

## 4.3 實驗二

### 4.3.1 實驗目的與設計

在同一k值下，在不同missing rate程度下所能夠找回近似skyline set的程度。

### 4.3.2 實驗方法

本實驗所使用的資料集k值最大範圍可以到17，故分別取三種不同k值分別做三次比較，以觀察k鄰近填補法、權重型k鄰近法以及本論文方法填補後的值所能夠找回原skyline set的程度。

x軸為缺失值佔整體資料集當中的比例，y軸為最填補所有缺失值之後，再分別跑同一支尋找skyline set的程式，並與缺失前的原skyline set做比較計算出相似程度。若越接近原skyline set則y軸的值越接近1.0。

### 4.3.3 實驗結果、分析與實驗結論

實驗二結果顯示出，本實驗方法雖然在某些k值不大情形下準確度會略差，如表4.2與圖4.4中，當k=5且缺失比率0.3到0.55之間，本論文方法準確度有下降，此時k鄰近填補法有機會擁有較好的填補效果是因為缺失程度不高下，k鄰近填補法還能夠以足夠的k與鄰近點計算平均後填回。

本論文的方法之中有採取採樣的機制，此機制在k值不大的時候且缺失比例不高時會比較容易啟動，但隨著缺失比率增加下，k值的增加所帶來的益處會越不明顯，而此時本論文中採樣的方式反而可以起到更大機會能有效的找到鄰近參考點。

從表4.1與4.2中可以看出k鄰近填補法大約從缺失比率30%時準確度就開始急遽下降，雖然中間可能有小幅度的上升，但可看出下降程度在75%以後下降幅度又更加嚴重，表4.3更是提前在缺失值達55%時就開始大幅下滑，這都顯示出一個現象，即k鄰近填補法所參考鄰近點的機制，對於缺失值很多的情形下抗性不高。

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **缺失比例** | **0.1** | **0.2** | **0.3** | **0.4** | **0.5** | **0.6** | **0.7** | **0.8** | **0.9** |
| **kNN** | **0.8** | **0.7** | **0.538** | **0.6** | **0.421** | **0.5** | **0.384** | **0.25** | **0.1** |
| **weighted kNN** | **0.8** | **0.7** | **0.538** | **0.6** | **0.421** | **0.5** | **0.285** | **0.25** | **0.1** |
| **proposed method** | **0.8** | **0.8** | **0.818** | **0.909** | **0.636** | **0.75** | **0.7** | **0.636** | **0.545** |

表4.1

圖4.3 k=1

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **缺失比例** | **0.1** | **0.2** | **0.3** | **0.4** | **0.5** | **0.6** | **0.7** | **0.8** | **0.9** |
| **kNN** | **0.9** | **0.666** | **0.666** | **1** | **0.625** | **0.533** | **0.6** | **0.454** | **0.181** |
| **weighted kNN** | **0.9** | **0.625** | **0.6** | **1** | **0.529** | **0.692** | **0.666** | **0.1** | **0.25** |
| **proposed method** | **0.9** | **0.833** | **0.909** | **0.818** | **0.5** | **0.8** | **0.6** | **0.5** | **0.363** |

表4.2

圖4.4 k=5

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **缺失比例** | **0.1** | **0.2** | **0.3** | **0.4** | **0.5** | **0.6** | **0.7** | **0.8** | **0.9** |
| **kNN** | **0.9** | **1** | **0.75** | **0.333** | **0.727** | **0.583** | **0.615** | **0.384** | **0.181** |
| **weighted kNN** | **0.9** | **0.833** | **0.818** | **0.4** | **0.277** | **0.529** | **0.428** | **0.352** | **0.181** |
| **proposed method** | **0.9** | **1** | **0.833** | **0.454** | **0.818** | **0.7** | **0.727** | **0.75** | **0.7** |

表4.3

圖4.5 k=13

## 4.4 實驗結果與分析

由實驗二中的圖可以知道無論是k鄰近填補法或是權重型k鄰近填補法，在k值分別為1、5與13時，隨著missing rate在整體資料集中升高，亦無法有很好的填補效果，經分析其原因有二 :

1. 在考慮最鄰近的k值作為參考該維度的值時，可能會遇到不足k個不為NaN的值，而傳統k鄰近填補法在遇到此種情況時，會選擇從缺不補，使得剩下的不足k個被參考鄰近點該維度值之權重無形中上升。

當缺失比例越高的時候，此狀況也就越趨明顯，最後k鄰近填補法所填補的新值雖然為平均，但也幾乎被簡化為單一值填補法的效果，最終效果如同只填補均值、眾數、或最大最小數結果一樣。

1. 遇到NaN值計算距離的機制，當兩兩資料點計算出距離時，若相對應的

而上述兩個原因亦無法藉權重法求加權平均數來彌補此一現象的缺陷，因此亦可看出即使採用權重型k鄰近填補法也不會有太好的填補效果，從此可以得知，挑選可參考性鄰近點的值在高缺失值比例下，其影響力遠比參考更多鄰近點來的大。

# 第五章 結論與未來方向

## 5.1 結論

### 5.1.1 學術貢獻與結論

在面臨需要考量使用者喜好的應用問題時，例如根據使用者使用習慣或評分的推薦系統，每個特徵欄位均需要被考慮的情形下，通常以 skyline algorithm 是最為常見的，而 skyline algorithm 又仰賴於輸入資料集的完整性(completeness)，但若是無法順利取得完整的資料集、或是具有缺失值使得所能取得完整資料筆數不足，在大部分情況下某一缺失值是無法與該特徵欄位其他非缺失值相比較的，此時 skyline algorithm 就無法針對每一筆資料的所有特徵欄位做比較，導致最終無法得出skyline set points，有鑑於缺失值存在的問題，本論文提出的方法基於原k鄰近填補法之上同時考慮缺失情形與計算距離方式進而決定一個權重向量weight vector w對其k個鄰近點分配以新的權重值來做為原缺失值欄位被填補的新值，如此一來不僅不會因刪除而導致喪失關鍵的資料特徵，也可以填補比原本k鄰近填補法更具有參考性的值，因此填補後對於 skyline set 的影響也是本論文所關注的重點之一，其中也會對不同填補法所跑出來的 skyline set points 結果做評估。

本論文方法係鑒於k鄰近填補法概念當中的考量鄰近點的基礎之上，與傳統k鄰近填補法最大的不同是在於傳統k鄰近填補法對於某缺失值其考量所有的鄰近點在該特徵欄位上只單純取算術平均後以填補新值，而本論文所提出的方法則是對其周邊鄰近點根據缺失程度取加權?平均，以此來改善原方法

本論文方法在 missing rate 介於 20% ~ 80% 所填補值對於最終 skyline set 的結果均具有不錯的填補效果，尤其是當該欄位缺失很集中或是缺失比例偏高的狀況下，其所填補後結果以 hamming distance 測量出來的精準度都不錯，表示本論文方法在填回去 skyline set points 中具有一定的效果。

## 5.2 未來工作與方向

### 5.2.1 方向與可探討的解

若是能夠透過資料集提供的資訊，綜合考量之後挑選出最有可能成為 skyline set points 的因素，考慮 skyline 與 non-skyline 佔其鄰近點比例也可能會對最終填補值的準確程度造成若干程度的影響，畢竟如果考慮的 skyline 的值越多，則具有該缺失值的點與在特徵 skyline set points 的相似程度可能越大

# 參考書目

## 一、中文部分

## 二、西文部分

說明：

1.在正文撰述過程所徵引的所有參考文獻資料均應編製成參考書目置於本文之後，另起一頁，頁次接續正文編碼。內容首分中、西文，中文在前，西文在後；次分圖書、期刊論文、網路資源及其他。

2.參考書目之格式依各學門習用者為準，或採用MLA、Chicago Manual、American Psychological Association……等格式。

# 附錄

說明：凡屬大量數據或冗長資料，不便刊載於正文者，如原始資料（Raw data）、訪問記錄或問卷等，均可編為附錄置於參考書目之後，另起一頁，頁次接續參考書目編碼。