<https://www.cnblogs.com/1zhk/p/4776027.html>

**1推導近似度(Deriving proximities)**

用於MDS分析的數據稱為**近似度**。**近似度**指出所調查客體(object)與總體(overall)之相似性或不相似性。MDS程序會尋找對象的空間配置，以使對象之間的距離盡可能接近它們的**近似度**。通常數據以方陣-**近似度**矩陣之形式排列。有2種主要的推導近似方法：直接方法&間接方法。

**1.1直接方法**

受試者可以為每對對象分配一個數值(不)相似值，或者依據它們的(不)相似度提供對的排名。 兩種方法都是收集**近似度**數據的直接方法。

**(不)相似等級**

一般在評級時，被要求以數字表達兩個對象的相似性，通常使用7級或9級。 在不相似量表中，低數字表示兩者間的強相似性，而高數字表示極不相似性。 為了分級，必須將所有可能的對像(objects)對呈現給參與者[總數為n(n-1)/ 2，其中n是對像數]。

然而這假設不相似關係是對稱的，因此每對中的順序無關緊要。 許多MDS程序也可以處理非對稱**近似度**數據。(例：a與b相比，b與b更相似； 例；在收集混亂數據時，可能會出現不對稱的**近似度**)。 但研究不對稱的鄰近關係會使要呈現的對數增加一倍。

直接評級的優點是數據可以立即用於MDS分析。 因此，既可以對每個參與者進行單獨調查，又可以基於**近似度**矩陣的平均值進行匯總分析。 相似度評定的缺點是，隨著對像數量的增加，成對比較的數量會迅速增長。

**排序任務**

有很多方法可以取得主題相似性的順序。排序方法的優點是對主題非常直觀，但是其中一些不允許對數據進行單獨分析。

Method 1：將每對對像寫在卡片上，然後讓參與者從相似性最低到最高的順序對卡片進行排序。

Method 2：將相似度低的對的卡片放到一堆中，將相似度較高的對的卡片放到下一堆中。依此類推。再將(不)相似性的數值分配給每個堆。

Method 3：在每個卡上只寫一個對象，並要求參與者將具有最相似對象的卡分類為一堆。 計算兩個對像在一堆中一起出現的次數，得出**近似度**矩陣。

**1.2間接方法**

用於**近似度**數據的間接方法不需要直接為**近似度**矩陣的元素分配數值。 而**近似度**矩陣是從其他度量導出的。(例：來自**混淆數據**或**相關矩陣**)。

**混淆數據**

當研究人員記錄受試者將一種刺激(stimulus)誤認為另一種刺激之次數時，就會產生混淆數據。考量一個字母表的字母通過揚聲器簡略地出現之實驗。主題的任務是識別字母。從數據中可以得出一個**近似度**矩陣：很少混淆的字母會得到較高的相異值，而經常混淆的字母會得到較低的相異值。

使用混亂數據的一個優點是可以在感知水平上判斷刺激的相似性，而無需進行過多的認知處理。 因此使用這種技術可能會揭示出非常基本的感知維度。 另一方面，混淆數據通常是不對稱的，並且不允許進行單獨的分析。 最值得注意的是，必須有一個很好的機會將一個物體與另一個物體混淆，這排除了使用這種方法研究可完全區分的刺激的可能性。

**相關矩陣**

MDS的另一個應用是將其用於可視化相關數據。 當物體在不同的尺度上測量並且測量值相互關聯時，就會形成一個關聯繫數矩陣。 即使只有幾個對象，這樣的矩陣也變得複雜，並且很難檢測相關模式。 MDS解決方案將對象繪製在地圖上，以便通過視覺檢查可以訪問它們的相關結構。

表1顯示了一個相關矩陣的示例：它列出了美國50個州收集的犯罪率之間的相關係數。僅從數據來看，不容易看出哪些犯罪率是相關的。圖1中的MDS表示簡化了很多任務。 圖中的距離對應於相關係數，因此，高相關性由較小的距離表示，反之亦然。 除了圖形表示之外，MDS分析還通過解釋MDS空間的軸來說明相關性：x軸可能解釋為“人與財產”，y軸解釋為“隱藏與街道”。

將MDS應用於相關數據可能比僅報告相關係數更生動地揭示對象之間的關係。該方法的缺點是需要根據其他測量結果來推導**近似度**。

另一種推導**近似度**的方法不需要這種測量。因此即使事先不知道所關注刺激的比例，尺寸或屬性，也可以進行MDS分析。實際上導出這些維度是分析的目的。

導出**近似度**數據的直接方法和間接方法均會產生**近似度**矩陣，該矩陣用作MDS程序的輸入。在許多實際應用中，直接要求參與者對對象(不)相似性的判斷是很直接的。 另一方面，間接方法可能非常適合於基本感知維度的研究，或者在已經存在研究對象的其他度量的情況下。

**2如何進行MDS?**

MDS分析的目的是找到對象的構圖，這些對象之(不)相似性的某些度量為已知。該等構圖應提供一些洞察力，使受測者可評估了解潛在未知維度的刺激項。當得出了**近似度**，數據收集就結束了，必須使用計算機程序確定MDS解決方案。

MDS分為「古典(計量)」與「非計量」。古典MDS假定數據(**近似度**矩陣)顯示出度量的性質(例:從地圖測量的距離)。故古典MDS空間中的距離會盡可能保留鄰近點之間隔和比例。

故非計量MDS僅假設**近似度**的順序有意義。

在非計量MDS構圖中，距離的順序盡可能反映**近似度**的順序，而間隔和比例的訊息則無關。

為了清楚理解MDS結果，簡介兩種MDS程序的基本機制(古典和非計量MDS)可能會有所幫助。

**2.1古典MDS**

考量以下問題：查看顯示許多城市的地圖，對其中一對間的距離感興趣。 通過則以尺測量它們的距離；另外亦可用數學方法(歐式距離Euclidean distance)。

考量相反的問題：僅具有距離，是否可以獲得地圖？古典MDS由Torgerson (1952)首次提出，旨在解決這一問題。 它假定距離為歐式距離。歐式距離通常是MDS空間的第一選擇。 但存在許多非歐式距離測度，這些測度僅限於非常具體的研究問題(參見Borg＆Groenen，1997)。在MDS的許多應用中，數據不是從地圖測得的距離，而是**近似度**數據。當將古典MDS應用於臨近點時，假定臨近點的行為類似於實際測得的距離。 這可能適用(例:用於從相關矩陣得出的數據)，但很少用於直接相似性評級。古典MDS的優點在於它提供了一種不需要迭代過程的分析法。

**古典MDS演算步驟**

古典MDS算法通常涉及一些線性代數。不熟悉這些概念的讀者也可以跳過下一段(參見Borg＆Groenen，1997，以進行更仔細的介紹)。

古典MDS算法依下列：坐標矩陣**X**可以通過特徵值分解由內積矩陣**B** = **XX’**導出。 藉由**近似度**矩陣**P**及矩陣**J = I–(1/n)E**可以解決構造**B**的問題。 此過程稱為*double centering*。 以下步驟總結古典MDS算法:

Step 1.建立平方近似度矩陣：**P(2)=[p2]**

Step 2.使用**J = I–(1/n)E**計算*double centering* : **B=(-1/2)J P(2)J** , **n**:objects數量

Step 3.求**B**的*m*個最大特徵值及其對應之特徵向量。

Step 4.從坐標矩陣**X =Em(Λm)1/ 2**導出**n**個對象的**m**維空間配置，其中**Em**是**m**個特徵向量的矩陣，**Λm**是**B**的**m**個特徵值的對角矩陣。

**2.2非計量MDS**

當使用MDS探索人類主體的感知空間時，將**近似度**行為類比為距離的假設可能過於嚴格。

為了克服這個問題，Shepard（1962）和Kruskal（1964a，b）開發了一種稱為非計量MDS。在非計量MDS中，僅以**近似度**的順序(ordinal)建構空間構圖。

計算**近似度**的單調轉換，產生縮放的**近似度(scaled proximities)**。最佳縮放**近似度**有時稱為差異性(*disparities*)**d’= f(p)**。

非計量MDS的問題是如何找到一個點的構圖，使最佳縮放**近似度**與點間距離的平方差最小化。

換句話說，讓**p**表示**近似度**向量(i.e**近似度**矩陣的上三角或下三角)，**f(p)**表示**p**的單調轉換，而**d**表示點距離； 然後必須找到坐標，以最小化**壓力**(*stress*)

MDS程序自動將壓力降至最低，以獲得MDS解決方案； 然存在許多不同的壓力版本。

**評判配合度**

應力的大小也可用於判斷MDS解答配合度的優劣：較小的應力值表示配合度高，而較高的應力值表示配合度不好。 Kruskal（1964a）提供了一些關於解答配合度應力值評定準則(表2)。

**注意：**這些簡單的準則很容易被濫用。 為了避免誤解，請牢記以下幾點：

* MDS文獻中有許多不同的應力公式。 本案原則僅適用於由(2)式計算的壓力，有時也稱為Stress1。
* 應力隨著維度數量的增加而減小。故二維解答始終比三維解答具有更大的壓力。

由於應力絕對值僅能模糊地指出配合度的優劣，因此通常有兩種附加的技術可用來判斷MDS解答的充分性：碎石圖和Shepard圖（參見Borg＆Groenen，1997； Hair ，Anderson，Tatham＆Black，1998）。

在碎石圖中，應力值與維度作圖。由於應力隨著維度的增加而單調減小，因此人們一直在尋找具有可接受應力的最小維度。碎石圖中的“肘部(elbow)”表示，更大的維度只會在應力值產生較小的改善。故最合適MDS模型之維度數量與碎石圖中肘部的維度數量一樣多。

Shepard圖顯示了**近似度**與構圖中點的距離間之關係。在此圖中，較少的分佈(spread)表示良好的配合。在非計量MDS中，Shepard圖中點的理想位置是單調遞增的線，描述了所謂的差異性，即最佳縮放**近似度**。

圖3顯示了範例性的碎石圖和Shepard圖。 碎石圖上的“肘部”表示三維MDS空間，而Shepard圖中很少的分佈(spread)表示該方法的配合度很高。

**非計量MDS演算基礎**

非計量MDS演算的核心是雙重優化過程。

首先必須找到最佳的單調轉換。 其次必須優化構圖中的點，以使它們的距離盡可能接近縮放的**近似度**。

非計量MDS算法的基本步驟如下：

1.隨機找出構圖中的點，例:通過從正態分佈採樣。

2.計算點之間的距離d。

3.找到近似度的最佳單調變換，以獲得最佳縮放近似度f(p)。

4.通過找到新的構圖中的點來最小化最佳縮放近似度和距離之間的壓力。

5.將壓力與標准(scree)進行比較。如壓力夠小，則退出算法，否則返回2。

**3.以MDS評級聲音品質**

以下的研究係使用不同類型MDS應用於分析聲音品質的例子。

數據由克里斯汀·施密德(Christian Schmid)於2002年夏天在奧爾堡大學的聲音品質研究部門(SQRU)所收集。該研究的目的之一是揭露和確定受試者在評估環境聲音時使用的維度。共有77名受試者參加了實驗。

通過耳機向他們展示了所有66對12種環境聲音。受試者的任務是將每兩種聲音的不相似程度從1（非常相似）到9（非常不相似）進行評分。所得的近似矩陣既可以進行個體分析，也可以進行合計分析。

**3.1個體分析**

對於個體分析，選擇使用非計量MDS模型，因為懷疑數據中是否包含序數信息。以個體37為例。

表3顯示了**近似度**矩陣，包括相異度等級。

有了12個刺激(stumuli)，就有可能在最多三個維度上尋找空間表現形式(參見第4節)。

圖4中的碎石圖可以幫助找到合適的維度。 顯然，一維表示是不夠的(應力> .30)。碎石圖沒有清晰的肘部，但是從一維更改為二維時，發生最大應力改變。故選擇了二維解決方案。 二維解的應力為0.156。 Shepard圖顯示了二維解的配合度。最佳配合度的點應位於單調遞增線上。Shepard圖中的分布顯示與理想配合度有一些偏差。但它被認為足夠小，可以進行進一步的分析。

在圖5(二維圖)中顯示個體37的**近似度**(?)。解釋此類MDS圖的一種策略是尋找客體(objects)群。例:“打字機”、“輪胎壓碎石”和“狂風”這三個客體似乎組成一群，因為它們彼此間的距離比其他任何聲音都近。另一種方法是取兩個非常遙遠的客體，然後嘗試對維度進行解釋。 在第二維度上，“投石入井”和“黃蜂”的聲音似乎非常不同，但在第一維度上，聲音卻相似。但這些維度的實質性解釋並不明顯。

**3.2合計分析**

為了執行合計MDS分析，通過計算每個cell的平均值來合併77個個體近似度矩陣。同樣，選擇具有歐式距離的非計量MDS模型來表示數據。二維解的應力為0.106；從一維變為二維時，應力發生最大的改變。構圖如圖6.。

如果還有可用刺激體(stimuli)的其他度量，則可以通過將維度與外部度量相關聯來搜索維度的經驗解釋。在本案，由另一個具有相同聲音和大致相同主體的實驗中，得出了不愉快的量表。(i.e每種聲音的不愉快值是已知的)。使用Spearman等級(rank)相關性將第一維度(x軸)上的值與不愉快程度相關聯，得出統計學上顯著的相關性69％。

粗略地講，圖6中沿x軸的變化的一半可以由聲音的不愉快來解釋。這一發現強調了一種不愉快的心理措施對於感知環境聲音的重要性。

**3.3個體差異量表**

個體差異量表(INDSCAL)[或:加權MDS]由Carrol＆Chang(1970)次提出。使用這種技術，有可能既代表共同的MDS空間中的刺激體，又代表個體的差異。

為了實現這一點，假設所有主體在評估客體時都使用相同維度，但是可以對這些維度施加個別權重。通過估計各個權重並繪製它們(在低維度的情況下)，可以偵測到不同的主體群。INDSCAL分析是所有主體的個別**近似度**矩陣輸入。

圖7顯示了應用於十二種環境聲音的INDSCAL程序之結果，以及上一節中介紹的77名參與者的樣本。MDS圖看起來類似於圖6中所示的MDS集合。在圖7的右側，主體權重的緊密聚集表明樣本相當均勻。

圖的左下角的受試者31、36、55和73可以被視為單獨的一 群；也就是說，與其他樣本相比，這些參與者在二維上的權重較小。

**結論**

本節旨在說明可能在聲音品質研究中使用的各種MDS技術。

(1)在評估環境聲音時，主體使用的維度大致相同

(2)評估維度之一(X軸)與聲音不愉快感高度相關。

**4.事前決定事項**

MDS需要研究人員具備一定的專業知識。與單變量統計方法不同，MDS分析的結果更多地取決於事先做出的決定。在數據收集階段，應該意識到要求相似性而不是不相似性評級可能會影響結果。也就是說，相似性判斷不能簡單地視為差異性判斷的“相反”。此外，必須分別在直接方法和間接方法以及對稱數據收集方法和非對稱數據收集方法之間做出決定(請參閱第1節)。

建立**近似度**矩陣的方式可能已經決定適當的MDS模型的選擇。 如果**近似度**使得實際數值幾乎沒有意義，並且排名順序被認為是唯一的相關信息，則應選擇非計量模型而不是計量模型。 此外，當分析的最重要目標是可視化數據中的結構時，建議使用歐幾里得距離。 非歐幾里得距離會掩蓋目視檢查的結果，但它們可能是研究有關對象感知空間的特定假設的有價值的工具。 最後，選擇的壓力測量類型將影響MDS表示。

顯然，MDS空間的維數將極大地影響解決方案。

先驗假設可能有助於選擇合適的數字。例:如果問題為是否可以一維尺度表示對象，那麼一維MDS解決方案將是最令人感興趣的。

要縮放的客體數量給出了進一步的指導原則(Borg＆Groenen，1997)：一個k維表示至少需要4k個客體，即二維表示至少需要八個客體。後驗，壓力量和可解釋的維數將提供有關選擇多少維度的更多信息。

另一個重要的決定是要執行的MDS分析的類型。

同樣，具體的研究問題可能會決定選擇。個體分析最準確地表示每個主體的數據。但通常人們對個體差異不感興趣，而對“平均”主體的感知空間感興趣。在這種情況下，合計分析更為合適。INDSCAL提供了常用的MDS空間和個體差異的表示。但這種情況下，可假設所有主體共享一個共同的感知空間，並且僅在其維度權重上有所不同。

最後，必須找到可以進行分析的適當的MDS軟件。 所有主要的統計軟件包，例如SPSS，SAS，Statistica和S-Plus，都可以執行MDS分析。 除了該獨立軟件之外，還存在專門為MDS設計的獨立軟件。 當前最常用的一種是Forrest Young的ALSCAL。 可從http://forrest.psych.unc.edu/research/alscal.html免費獲得。 然而，隨著軟件的選擇，在較早階段的一些決定可能是預定的，例如。 壓力測量的類型或分析的類型。 在購買或使用MDS軟件時，應再次確保它不僅可以執行經典MDS，而且還可以執行非度量MDS。 這適用於任何上述軟件。 圖8顯示了MDS分析之前的階段以及研究人員必須選擇的一些選項。