

# 港澳特區政府施政報告的比較研究 (2000-2010) \*

伍向豪\*\*、溫雅筠\*\*

## 一、緒言

本文的研究動機主要探討香港特區和澳門特區政府所採取的政策方針。由於澳門特區的施政具有行政主導的特色和香港不盡相同，因此從港澳特區特首施政報告可以得知港澳特區政府施政上所呈現的特性，對比較港澳政府施政方針具有幫助。因此本論文主要以文本採礦法分析港澳特區政府2000年至2010年的施政報告，以歸納澳門政府在施政上所偏重的政策方向。

本文架構大致安排如下：第二部份首先介紹本文的研究方法，再介紹文本採礦方法（Text Mining）分析澳門政府2000年至2009年的施政報告的詳細過程，藉以瞭解澳門特區政府從1999年底回歸中國至今的施政方向。第三部分則就港澳特區政府施政報告的文本挖掘結果進行，最後提出本文結論。

## 二、研究方法

### （一）研究方法概述

本文認為，要瞭解港澳特區政府所面臨的政策環境，需從港澳特區政府每年的施政報告著手，因此本文在研究方法部分主要介紹文本採礦分析，以文本採礦方法得出2000年至2010年施政報告中的關鍵

---

\* 澳門科技大學行政與管理學院助理教授。

\* 澳門科技大學行政與管理學院助理教授。

\*\* 臺灣屏東科技大學管理學碩士。

字，藉以瞭解十年來港澳政府主要的施政主軸，並分析施政主軸的優缺點。

## （二）文本挖掘法概述

由於文字資料的數量和類型十分龐大，為求掌握這些文字資料中有用的知識，資料採礦（Data Mining）的相關技術常使用在文字資料的知識萃取用途上，這種使用資料採礦方法於文字資料的特殊技術，一般稱為文本採礦（Text Mining）。文本採礦（文字探勘、文字挖掘，Text Mining）的應用主要在網路資訊管理、生物資訊和客戶關係管理等領域。在其他領域的應用則以圖書資訊、文學研究領域為主。例如Clement以文本採礦方法就Gertrude Stein所著的美國人的型塑（The Making of Americans）進行文本分析得出和文學批判（Literature Critics）學者迥然不同的結論。在其他領域中的應用則以Trappey et.al.就專利說明書的文本採礦研究為特殊應用範例。

至於在公共政策領域，文本採礦方法的應用目前較少。臺灣“國科會”政治學門“熱門及前瞻學術研究議題調查”五人工作小組曾於2006年針對質與量結合的方法論及研究方法提出將文本採礦方法應用在以質化方法建立的文字資料庫，例如立法院公報、政治人物之文告宣示、以及民調焦點團體訪談之文字紀錄稿等。

本文以文本採礦方法以2000年至2010年港澳特區政府施政報告作分析主體，萃取有用的資訊（關鍵字），藉以探討目前政府施政的主軸，對澳門博彩業的行銷策略的研擬和未來澳門政府所應採取的施政方向有參考作用。同時，本文也是文本採礦方法應用在公共管理領域的新嘗試。

## （三）文本挖掘法的應用

本文以R開源軟件（Open Source Software），以該軟件中的文字

挖掘插件 (Text Mining Package) 分析港澳特區政府2000年至2010年的政府施政報告。文本挖掘分析程序如下：

### 1、文本檔案語法分析 (Flat Document Parser)

首先就2000年至2010年英文版的港澳特區施政報告各按年份儲存在.txt檔案中，建立檔案夾，然後調用文本挖掘插件開始進行分析。

### 2、字詞原型演算法 (Word Stemmer)

本程式以Kuhlem詞語原型演算法將文本抽出的詞語轉為詞語的原型。

### 3、標點符號去除 (Punctuation Erasure)

本程式去除所有含有標點符號的詞語。

### 4、停止字串過濾 (Stop Word Filter)

本程式將從文本抽取的詞語將停止字串 (如about、above、across等等) 去除。停止字串清單系參照<http://www.lextek.com/manuals/onix/stopwords1.html>。

### 5、數字篩檢程式 (Number Filter)

本程式的目的在於去掉所有包含數位的詞語。

### 6、詞頻—逆向文件頻率 (Term Frequency—Inverse Document Frequency)

是一種用於資訊檢索與文本挖掘的常用加權技術。TF-IDF是一種統計方法，用以評估一字詞對於一個文件集或一個語料庫中的其中一份文件的重要程度。字詞的重要性隨著它在文件中出現的次數成正比增加，但同時會隨著它在語料庫中出現的頻率成反比下降。TF-IDF加

權的各種形式常被搜索引擎應用，作為文件與用戶查詢之間相關程度的度量或評級。逆向文件頻率衡量詞語在文本中的普遍重要性，某一特定詞語的IDF，可以由總文件數目除以包含該詞語的文件數目，再將得到的商數取對數，即：

$$\text{idf}(t) = \log(1 + (f(D) / f(d, t))) \quad (1)$$

其中 $f(d)$ 是文件總數， $f(d, t)$ 是文件中包含詞語 $t$ 的數目。

#### (四) 文本挖掘分析結果

為求能夠瞭解施政報告裡的主要意涵，本文將這些關鍵字的詞頻—逆向文件頻率使用主成份分析方法進行分析。本文使用R軟件中的rattle插件以奇異值分解（SVD; single value decomposition）主成份分析法進行分析。奇異值分解主成份分析法係將所欲分析的資料分解成如下的主成份，這裡以兩個主成份來舉例（Joliffe, 2002）：

$$\chi_{p1} = \mu_1 + \Lambda_1 z + \Gamma_1 y_1 + e_1 \quad (2)$$

$$\chi_{p2} = \mu_2 + \Lambda_2 z + \Gamma_2 y_2 + e_2 \quad (3)$$

在（2）和（3）式中， $\mu_1$ 和 $\mu_2$ 為平均數的向量， $\Lambda_1$ ， $\Lambda_2$ ， $\Gamma_1$ ， $\Gamma_2$ 是係數矩陣， $z$ 是 $\chi_{p1}$ 和 $\chi_{p2}$ 的隱性變數（latent variables）向量， $e_1$ ， $e_2$ 是殘差向量。

##### 1、香港特區2000年至2010年施政報告文本挖掘分析結果

本文以香港特區2000年至2010年英文版施政報告做文本挖掘分析，按照上述的分析程序首先按照詞頻大小得到services（服務）、business（商業）、community（社區）、development（發展）、economic（經濟）、education（教育）、enhance（提升）、financial（金融）、mainland（內地）、market（市場）、measures（措施）、people（人民）、policy（政策）、promote（促進）、provide（提供）、public（公共）、scheme（階段）、social（社會），以及support（支持）等關鍵字。同時為求能夠更合理的呈現關鍵字

的全貌，本文在分析過程中將施政報告中常出現的特區（special administrative government）裡的三個英文字不納入關鍵字中。

奇異值主成份分析法的分析結果如圖1、圖2和表1：

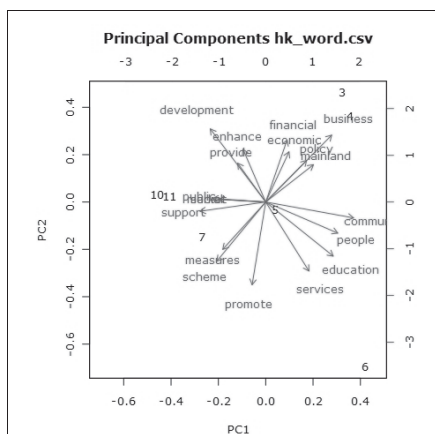


圖1：奇異值主成份分析圖

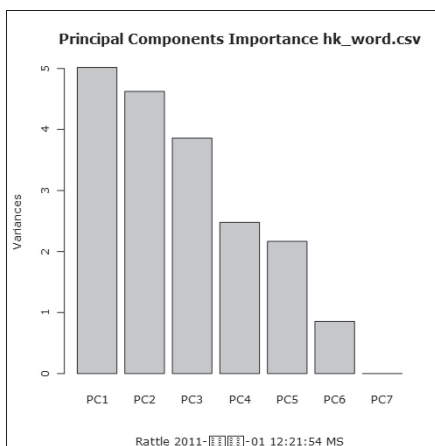


圖2：各主成份變異數圖

本文從香港特區2000-2010年施政報告關鍵字的奇異值主成份分析圖發現2002和2003年的施政報告的內容屬性介於經濟（economic）

和商業（business）兩個關鍵字之間。2004年施政報告的內容屬性介於服務（services）和人民（people）之間。2005年施政報告的內容屬性則介於服務（services）和教育（education）之間。2006年施政報告的內容屬性則介於措施（measures）和支持（support）之間。2009和2010年的施政報告的內容屬性則介於公眾（public）和支持（support）之間。

本文根據香港特區施政報告關鍵字主成份變異數分析圖，得知前四個主成份佔有較大的份額。

**表1：香港特區2000-2010年  
施政報告文本挖掘主成份分析結果**

	PC1	PC2	PC3	PC4
services	0.195	-0.324	0.024	0.216
business	0.298	0.314	-0.081	-0.164
community	0.396	-0.074	-0.035	0.016
development	-0.250	0.343	0.043	0.112
economic	0.105	0.235	0.286	0.083
education	0.304	-0.254	0.084	0.178
enhance	-0.102	0.248	0.197	0.182
financial	0.096	0.291	-0.045	0.448
mainland	0.215	0.175	0.239	-0.144
market	-0.212	0.012	-0.400	0.137
measures	-0.193	-0.222	0.305	-0.283
people	0.325	-0.146	-0.135	-0.317
policy	0.185	0.195	-0.171	-0.417
promote	-0.061	-0.389	0.030	0.099
provide	-0.126	0.182	-0.320	-0.353
public	-0.239	0.016	0.271	-0.262
scheme	-0.220	-0.278	-0.229	-0.114
social	-0.212	0.012	-0.400	0.137

	PC1	PC2	PC3	PC4
support	-0.296	-0.043	0.331	-0.096
各主成分變異數累計百分比	26.4	50.7	71.0	84.1

綜合上述可以發現這些主要關鍵字可以劃分成四個主成份。按照每個關鍵字在每個主成份裡的係數將每個主成份裡主要的關鍵字歸納如下：

第一個主成份（PC1）：community、ducation、people，因此可以歸納為社區和教育；

第二個主成份（PC2）：business、development、enhance、policy、provide，因此可以歸納為商業發展政策；

第三個主成份（PC3）：economic、mainland、measures、public、support，因此可以歸納為和內地的經濟合作與支持公眾的措施；

第四個主成份（PC4）：services、financial、market、promote、social、scheme，因此可以歸納為促進社會服務與金融市場範疇。

在本文中，歸納每個主成份中所屬的關鍵字係以每個關鍵字在每個主成份中的係數為標準，係數最大的即歸納至所屬的主成份中。

#### 1、澳門特區2000年至2010年施政報告文本挖掘分析結果

本文另以澳門特區2000年至2010年英文版施政報告做文本挖掘分析，按照上述的分析程序首先按照詞頻大小得到society（社會）、citizen（公民）、community（社區）、continue（繼續）、cooperation（合作）、development（發展）、economic（經濟）、education（教育）、efforts（努力）、improve（改善）、local（地方）、people（人民）、policy（政策）、promote（促進）、public（公共）、quality（品質）、resources（資源）、service（服務）、support（支持），以及system（體系）等關鍵字。同時為求能夠更合理的呈現關鍵字的全貌，本文在分析過程中將施政報告中常出現的特區（special administrative government）裡的三個英文字不納入關鍵字中。

奇異值主成份分析法的分析結果如圖3、圖4和表2：

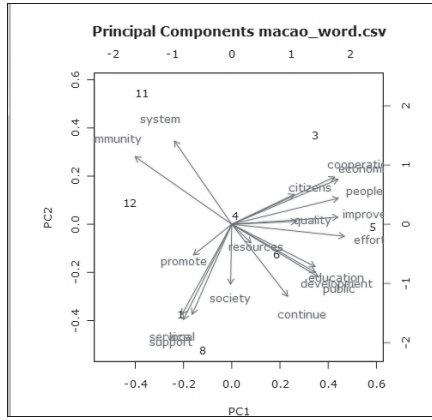


圖3：奇異值主成份分析圖

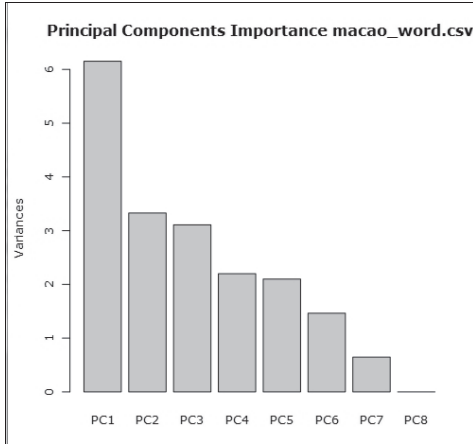


圖4：各主成份變異數圖

表2：澳門特區2000-2010年  
施政報告文本挖掘主成份分析結果

	PC1	PC2	PC3	PC4
society	-0.003	-0.245	-0.155	0.139
citizens	0.187	0.122	-0.382	0.099



	PC1	PC2	PC3	PC4
community	-0.290	0.276	-0.103	-0.164
continue	0.168	-0.294	0.039	0.157
cooperation	0.309	0.193	-0.267	0.120
development	0.251	-0.196	0.163	-0.383
economic	0.319	0.184	-0.125	-0.118
education	0.250	-0.173	-0.055	-0.101
efforts	0.338	-0.049	-0.126	-0.085
improve	0.319	0.029	0.036	-0.032
local	-0.118	-0.368	-0.085	-0.405
people	0.320	0.106	0.001	-0.338
promote	-0.115	-0.125	-0.442	-0.253
public	0.259	-0.214	0.229	0.029
quality	0.194	0.011	0.464	0.150
resources	0.057	-0.075	-0.418	0.277
service	-0.149	-0.368	-0.073	-0.317
support	-0.145	-0.389	0.053	0.324
system	-0.172	0.340	0.169	-0.273
各主成份變異數累計百分比	32.3	49.9	66.2	77.8

綜合上述，可以發現這些主要關鍵字可以劃分成四個主成份。按照每個關鍵字在每個主成份裡的係數將每個主成份裡主要的關鍵字歸納如下：

第一個主成份（PC1）：citizens、continue、cooperation、development、economic、education、efforts、improve、people、public，因此可以將第一個主成份歸納為：改進公眾教育和促進經濟發展及合作；

第二個主成份（PC2）：community、system，因此可以將第二個主成份歸納為社區體系；

第三個主成份（PC3）：quality，因此可以將第三個主成份歸納為質素；

第四個主成份（PC4）：society、resources、support，因此可以將第四個主成份歸納為社會支持和資源。

### 三、研究結果分析

按照上述的分析結果，本文就港澳特區施政報告主要關鍵字的主成份做比較：

（一）在香港特區施政報告主要關鍵字的主成份中，社區和教育為第一主成份，而商業發展政策為第二主成份。而澳門特區施政報告主要關鍵字的主成份中，改進公眾教育和促進經濟發展及合作為第一主成份，而社區體系為第二主成份。因此可以發現香港特區的施政方向係以社區及教育政策為第一方向，第二方向則為促進商業發展。而澳門特區的施政方向則以兼顧教育和經濟發展為首要方向，而社區政策為第二方向。本文認為，從港澳特區的施政方向來看，基本上符合新公共管理理論所強調的“以社區參與為本”的理論核心，而且亦強調經濟發展。

（二）由港澳特區施政報告中主要關鍵字的主成份分析可以得知香港特區施政報告關鍵字的第三主成份為與內地的經濟合作與支持公眾的措施，其變異數所佔的百分比為20.3%，顯見和內地合作發展經濟是香港特區施政報告裡的一個重點。而mainland（內地）這個主要關鍵字在澳門特區的施政報告中並沒有出現，這表示在香港特區的施政報告中，和內地的經濟合作是重要的核心之一，而在澳門特區的施政報告中並沒有特別強調。

（三）從香港特區施政報告關鍵字的四個主成份中，可以發現第二、第三及第四個主成份都和經濟、商業和金融市場有密切關係，顯示香港特區政府將施政主軸大部分都放在經濟發展上，而澳門特區的施政報告主軸雖然從關鍵字的四個主成份中做觀察，經濟發展與合作雖然屬於第一主成份，但第二個至第四個主成份和社會與社區政策有

關，因此本文認為澳門特區政府的施政除了經濟發展和教育以外，社會與社區政策是重要的施政主軸。這顯示兩地政府在施政主軸上的不同，也反映了兩地在發展上的相異性。

#### 四、結論

本文以文本挖掘法分析港澳兩地政府的施政報告，發現兩地政府在施政主軸上有明顯的不同。香港特區政府的施政主軸雖然主要偏重社區及教育政策，但是一大部分也側重金融產業及經濟發展，而澳門特區政府雖然主要偏重於教育和經濟發展，但也側重於社區及社會政策上。但值得一提的是，香港特區政府在施政上重視和內地在經濟上的合作，但澳門特區政府在施政報告上較未特別提及此點，也值得進行後續研究。

