

# 應用隱藏式馬可夫模型於口述對話系統之研究

顏國郎 吳宗憲 林建良

國立成功大學資訊工程研究所

Email: {yangl, chwu, lincl}@server2.iie.ncku.edu.tw

Fax : (06)2747076

## 摘要

本文目的在建立一應用於航空訂票及查詢系統中之對話系統。在論文中，首先是蒐集查詢及訂票對話的語料，並分析對話中的句子，推導及歸納輸入語句的語意。在我們的系統中，我們建立一包含語意的隱藏式馬可夫模型，此一模型將語音辨識後所產生的候選句作一評分後，刪除一些不可能的句子，然後再由剩下的找出語意分數最高的句子，作為語意的類別並輸出適當之句子，再經由對話管理員針對每一意圖作出適當之回應，以完成顧客查詢及訂位之目的。

## 一、簡介

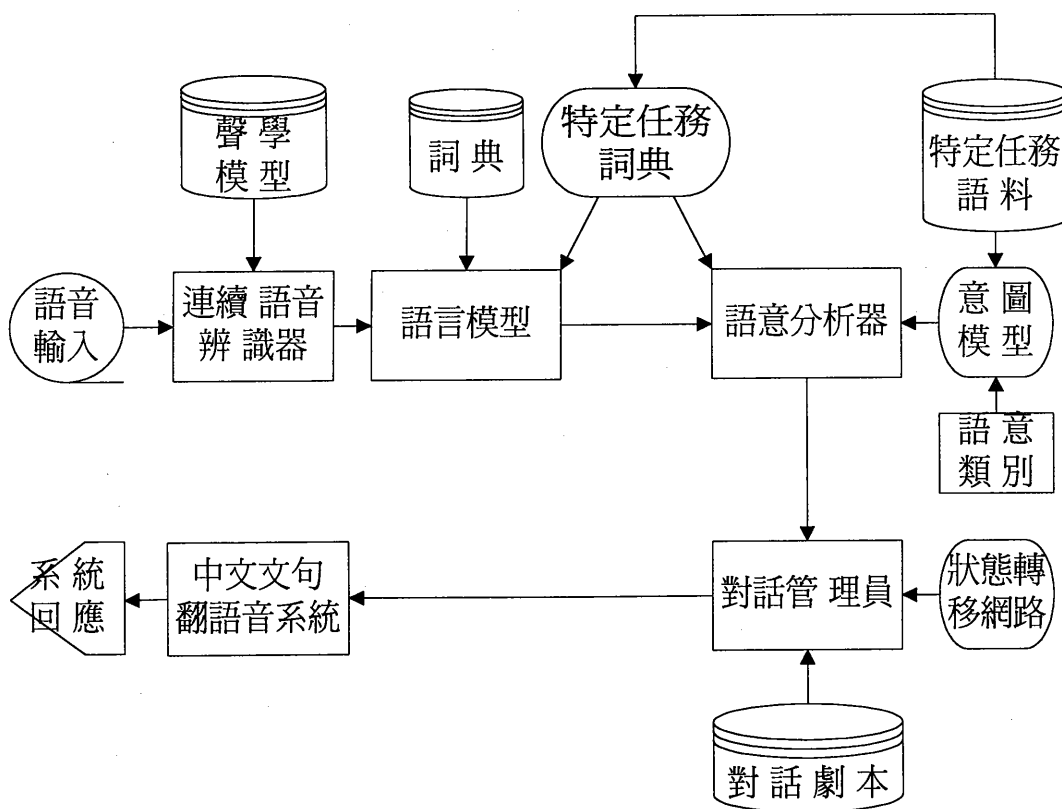
在過去幾年中，語音辨識技術已漸趨成熟，因此，口述語言對話系統也漸漸走向應用階段。例如資料庫查詢、客戶服務系統、自動總機系統、教學系統及機器控制等，均已有雛形系統相繼建立。在未來下一代電腦之基本功能上，電腦界已將對話口述語言理解列為一重要目標。然而對於無經驗之使用者與電腦間利用自然口述語言之對談，則仍存在一些問題。例如：大量詞彙之即時處理、對於不同口音使用者、自然對話處理之強健度以及人機對話中語意溝通之能力等，仍是大家正努力研究解決方法之主題。

由於完全理解人類的語言仍是相當困難，故目前的研究大多局限在其中的一個領域中。近年來已有很多項系統發表，如 MIT 之市區導覽系統、氣象資料即時查詢、車輛銷售導覽系統等[1][2]，其他如 Bennacef 及 Seide 等亦均有相關系統發表[3][4]，台大發展出銀行電話查詢系統[5]，中華電信研究所提出電話查詢服務系統[6]，而工

研院則發展出氣象查詢系統[7]，在成大則已發展出餐廳及電話查詢等系統[8][9][10]。近年來由於航空工業蓬勃發展，搭乘飛機已漸成一必須之交通方式，但由於搭機人數與日漸增，且班機時刻常常更動，因此，提供一良好之自動化航空資訊服務系統是大家所深切企盼的，所以，本論文乃針對航空資訊查詢及訂位之主題作一研究，以建立一對話模型。下面首先介紹的是系統架構，此部分著重於對話模型之研究，接著是對話語料之分析，我們說明對話語料如何分析取得結果，接下來是語意模型之建立，在此，我們採用隱藏式馬可夫模組(HMM)的觀念來建立語意模型，作為意圖的判斷依據。在對話模型，說明我們在對話中的對策與回應的方式。然後是一些實驗的結果。最後我們做一個結論與討論。

## 二、系統架構

在這章節中，我們介紹此一航空查詢及訂位系統之系統架構，其方塊圖如圖一所示。此系統可分為連續語音辨識器、對話模型和文句翻語音系統三大部分：



(圖一)系統架構圖

在論文中，我們針對理解部分建立一對話模型，而對於連續語音辨識及文句翻語音兩系統，將採用本實驗室自行開發之音中仙系統作測試。而對話模型共包含下列三個模組：

#### (一)、語言模型

在語言模型中，我們將建立一特定任務詞典，此詞典乃根據語料庫之資料而建立，在此語言模型中，首先接受語音辨識器輸出之 syllable lattice，而後根據特定任務詞典將其轉成 word lattice 並送至任務相關語意分析器作意圖之判定。

#### (二)、任務相關語意分析器

在語意分析中，先根據某一特定任務蒐集之關鍵詞建構詞段類別 (Word Segment Class) 而後針對可能出現詞段類別之組合建立一隱藏式馬可夫模型 (Hidden Markov Model, HMM)，每一個 HMM 代表一個詞段類別序列或意圖 (intention)，對於前述語言模型之輸出，選擇一最佳之意圖表示，而後逐一填入語意欄位，並判斷語意是否完整，否則，則透過對話管理員繼續詢問。

#### (三)、對話管理員

由於語音理解必須考慮語句前後之語意，因此，在對話管理員中，我們將對話記錄利用一狀態轉移網路表示，利用此一網路選擇所需之對話行為 (dialogue act)，並預測使用者可能接續之回應。

### 三、對話語料之分析

蒐集的對話語料可分為兩種：一種是語音格式，即錄製在錄音帶上的聲音，可用在辨識系統的訓練跟測試方面；一種是文字格式，便是將語音資料轉換為文字，用來分析對話句型、句子的語意、句子的文法等等，這部份的工作是屬於比較瑣碎的項目，所轉換出來的文字格式對話資料，分別以使用者與系統所說的話，分別一句句標示出來。我們分析所取到的語料，將每個訂票對話分成三個部分，如圖二所示：

(一)、問候語：主要是顧客與訂票人員間彼此間開始的問候語句。

(二)、資訊交換：其中分為

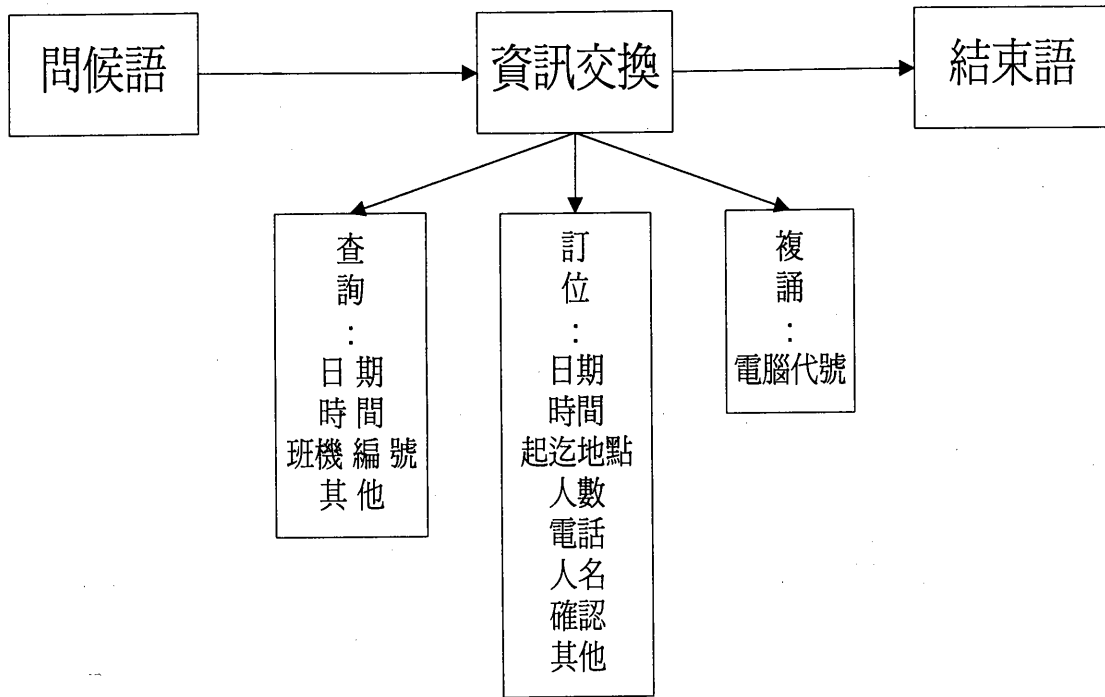
1、查詢：即使用者查詢班機日期、時間及編號。

2、訂位：為使用者針對系統所提的問題所做出的回答，或使用者本身提出訂位

資訊如班機之日期、時間、地點、人數、上下午、人名、確認及電話。

3、複誦：如電腦代號

(三)、結束語：訂位結束或查詢完成之對話語，如謝謝、再見。



(圖二) 訂票對話三大部分圖

針對上述之對話，我們利用對話語料庫建立一特定任務詞典，收錄其使用之詞彙及出現機率。另外在本系統中，有一些關鍵詞在語意上均頗相似，為使電腦方便處理我們將其作分類，因此，可蒐集到之特定任務詞根據其語意分成 21 種詞段類別(Word Segment Class)

- (1) 疑問詞：嗎，呢..
- (2) 寒暄語：你好、喂..
- (3) 詢問詞：請問、有沒有，幾點有..
- (4) 目標詞：班機、班次、機票..
- (5) From：從、由、自..
- (6) To：到、飛、回、往..
- (7) 肯定詞：是、對、好、可以..
- (8) 否定詞：除了、不是、不要..



- (9) 時間前約略語：大約、靠近、最早..
- (10) 時間後約略語：左右、以前、以後..
- (11) 動作詞：我要(訂)、我想(訂)、訂(位、票)..
- (12) 時間：早上、x點x分..
- (13) 班次：第x班、x班次..
- (14) 地點：台北、高雄、那邊..
- (15) 日期：x月x號、星期(禮拜)x、今天..
- (16) 人數：位、張..
- (17) Filler：哦、嗯、的、那個、怎麼..
- (18) 複誦：代號、英文字、數字..
- (19) 數字：1-59
- (20) 人名：李登輝、連戰..
- (21) Goodbye：：再見、謝謝、拜..

根據上述分類方式，每一輸入語句將可由 word lattice 轉成 class lattice，再送入語意分析器選擇最佳之語意序列。

#### 四、語意模型之建立

在語意模型方面，我們採用隱藏式馬可夫模組(HMM)的觀念來建立語意模型。建立的步驟分為蒐集關鍵字、關鍵字分類與訂定語意分析模組三個階段，以下分別簡述之：

1. 蒐集關鍵詞：這個階段主要的工作在於把所有蒐集到的語料，經過適當的斷詞處理，把它分為不重複且重要的詞，我們就稱此詞為關鍵詞。也就是把一個句子看成有順序的關鍵詞的組合。在電腦與人工的配合之下，我們採用半自動化的處理，對所有收集的語料共蒐集到 200 多個關鍵詞。
2. 關鍵字分類：這個階段主要對關鍵詞的詞段作分類，也就是 Word Segment Class。分類主要的用意是在作分析語意的走勢，對其作語意評估的依據。我們利用詞的前後相關來作為分類依據，也就是說相同語意的關鍵詞會有相同的前後相關詞。
3. 訂定語意分析模組：利用上面所分出來的關鍵詞類別，訂立語意分析模組是

這個階段的工作。我們所訂立的模組是採用隱藏式馬可夫模組，每一個隱藏式馬可夫模組代表一意圖(intention)，目前我們將其分成四種意圖，分別為 1. 問候語、2. 查詢、3. 訂位及 4. 結束語，而其他不屬於此四種意圖者則歸類至其他意圖，但不訓練 HMM。關鍵詞的類別是代表隱藏式馬可夫模型的狀態(state)，而在每個關鍵詞類別的關鍵詞出現的機率，代表隱藏式馬可夫模型的觀測機率(observation probability)，每一個狀態轉移即代表詞段類別之轉移方式。由這樣的觀念與動機，我們可以更精細的定義意圖 HMM 的基本要素如下：

離散觀測的意圖 HMM 中之成分定義如下：

- (1)  $N$ ：模組的狀態個數，而每個狀態代表一個詞段類別，所有狀態為  $\{1, 2, \dots, N\}$ ，對於第  $l$  個輸入關鍵詞片語，其狀態表示為  $S_l$
- (2)  $M$ ：代表每個狀態不同觀測符號的個數，每個觀測符號對應每個任務中的關鍵詞片語，我們可以把觀測符號表示為

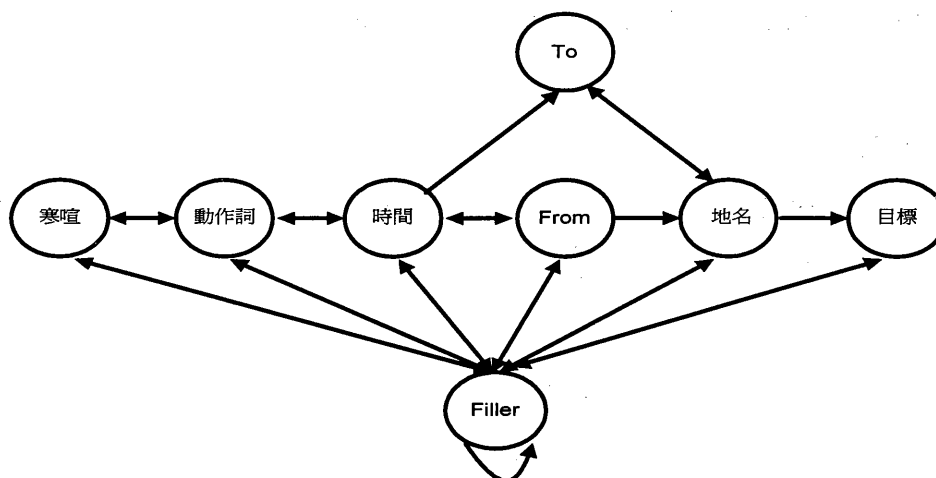
$$V = \{V_1, V_2, \dots, V_M\}$$

- (3) 狀態轉移機率  $a_{ij} = P(S_{l+1} = j | S_l = i), 1 \leq i, j \leq N$

代表狀態  $i$  轉移到狀態  $j$  的轉移機率

- (4) 觀測機率  $b_j(k) = P(O_l = V_k | S_l = j), 1 \leq k \leq M$  定義為狀態  $j$  的觀測機率
- (5) 初始狀態為  $\pi_i = P[S_1 = i], 1 \leq i \leq N$

我們所使用的 HMM 模組是採用比較寬鬆的轉移模組，即任何一個狀態均可以轉移到任何一個狀態，這個轉移模組可由訓練語料調整其轉移的方式而得到一個考量到 N-gram 的 HMM 轉移模組，圖三所示為一個訂位意圖所簡化之 HMM：



(圖三)訂位意圖之簡化 HMM

在語意辨識方面，對於一輸入語音  $U$ ，其對應片語序列之分數可以用下列的方程式決定：

$$S(PS_k | U) = \max_{1 \leq h \leq H} [\log P_h(WS_k | U)] + \alpha \sum_l \log P(ph_l^k | ph_{l-1}^k)$$

其中  $PS_k$  為第  $k$  個片語序列， $H$  為意圖 HMM 之個數。 $ph_l^k$  為第  $k$  個片語序列中的第  $l$  個片語。 $P(ph_l^k | ph_{l-1}^k)$  為片語的 bigram 機率。 $P_h(WS_k | U)$  表示對於一輸入語音  $U$ ，其所對應之第  $k$  個詞段類別序列  $WS_k$ ，經過第  $h$  個意圖 HMM 所得到之機率，其可用下式表示：

$$P_h(WS_k | U) = \max_{1 \leq i \leq N} \delta_L^{k,h}(i)$$

$$\delta_L^{k,h}(i) = \max_{S_1 S_2 \dots S_{l-1}} P[S_1 S_2 \dots S_l = i, o_1 o_2 \dots o_l | \lambda_h]$$

其中  $N$  為狀態數， $\delta_L^{k,h}(i)$  表示對於一觀測序列  $O = [o_1 o_2 \dots o_L]$ ，於第  $h$  個 HMM 中，利用 Viterbi 演算法，執行至第  $l$  個片語時所求得之最佳機率。

在 Viterbi 的程序中，觀測機率可用下列來計算：

$$b_j(O_l) = P(O_l | S_l = j) * PhS(ph_l^k)$$

其中  $PhS(ph_l^k)$  表示在第  $k$  個片語序列  $K$  中第  $l$  個的片語正規化的語音分數和片語相似度。

舉例子說明：若輸入對話為：你好，請幫我訂中午十二點台北的飛機。經過上面模組，會做出最適當的語意分析，並輸出最大可能的機率。其最佳路徑之詞段類別序列如下：

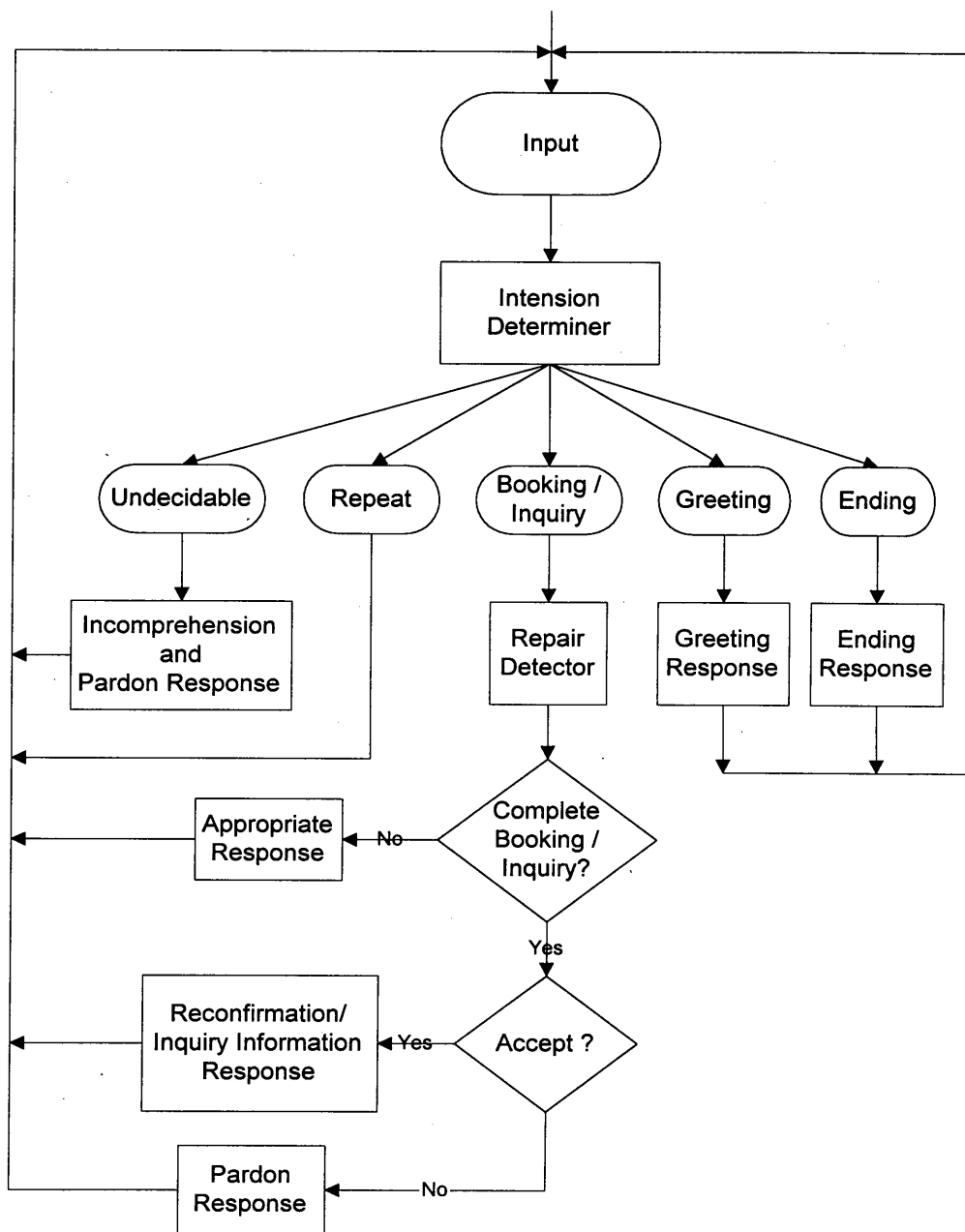
寒暄語→動作詞→時間→地點→Filler→目標

## 五、對話模型

在對話模型方面，我們將使用者的語意分成下列幾種：

- 一、問候：顧客對訂票人員的問候語句。如你好、是遠東嗎..等等，若知道使用者純粹是問候語，那我們便回應問候的語句及所需之服務。
- 二、訂票與查詢：要完成一個訂票或查詢過程中、都必須先有日期、起迄點、及時間這三個 Semantic Slot，所以將這兩種動作合在一起處理。若我們知道這是個訂票或查詢的動作，首先我們須從使用者的語句中擷取資料，因為我們不知使用者何時有講到那一個資料或一次講了那幾項，所以當知道使用者是一個訂位或查詢的過程時，我們必須分別對日期、起迄點、時間、人數去偵測並擷取出我們所須要的 Semantic Slots。我們不能因現在已有(時間)這個 Slot 的資料便不去偵測使用者的語句中有沒有這個資訊，因為使用者可能要更改已說過的資料，所以我們加了一個 Repair Detector 來偵測使用者是否有要更改 Slot 的動作。
- 三、重複：使用者最常重複系統的話是電腦代號，系統應經由使用者重複的同時檢驗使用者是否有得到正確資料
- 四、結束語：通常在對話的尾端不外乎是「拜(拜)」、「好(謝謝)你」等句子，代表使用者可能即將結束對話，此時針對 Slot 的情況給予適當的回應。
- 五、其他：一些無法處理的句子，系統告知無法了解並給予上次所提示的句子。

系統的對話模型架構圖如圖四所示：



(圖四)對話模型架構圖

在對話策略上我們採取以下的幾種方式：

1. 混合式交談：系統通常導引使用者回答飛機訂位所須要的資料，相對的，使用者也可以主動的去詢問所須要的資訊，例如：較接近下午三的飛機有那幾班。
2. 重複確認：為了確認系統所取到的資訊是正確的，每當從使用者的回答取得有關訂位所須的重要資料(Semantic Slot)，便複誦給使用者得知，且在對話將完成時複誦所有的資料給使用者做最後的確認。

3. 修復：若使用者查覺系統得到不正確的資料時，便可馬上糾正，以便與系統能做正確的溝通。
4. 覆蓋：假如使用者更正系統所回答的 Semantic Slot 那麼舊的 Semantic Slot 必須被修改，因此系統的對話須架構在新的 Slot 上做出適當的回應。

在系統回應的語句上，我們分成下面幾種類型：

- 一、問候：在對話的開始的回應句，如”你好”，”復興航空你好”。
- 二、詢問：當有 Semantic Slot 的資料未取得時，詢問使用者所須的相關資料。
- 三、選擇：當使用者詢問系統時，系統給予相關的資料且要求使用者輸入更多的資訊，如：

使用者：請問有沒有下午三點左右的班機

系統：較接近的有，二點四十分與三點三十分，請問你要那一班

- 四、重複：假如系統根據所辨識出來的句子，無法做出適當的回應，則系統告知使用者無法理解的資訊且要求使用者再輸入一次。
- 五、失敗：若系統超過三次無法做出適當的回應，則轉由人工處理。

在互動式的系統中，系統必須不斷地與使用者溝通，直到完整地處理完使用者的需求為止。而在訂位意圖之對話中，什麼時候才能算是對話的終結呢？我們定義，當使用者訂票的內容包含了下列七項語意類別時，我們就可以說這位使用者完成了一次的訂票流程：

---

[日期]、[時間]、[起點]、[終點]、[姓名]

[電話號碼]、[人數]

---

在對話過程中，系統要決定到目前為止，語意類別還缺了那幾項，且根據它的優先順序來決定回應語句，例如：要詢問[時間]時，必須先成[日期]與[起點、終點]這幾個語意類別。如此這個流程一直到產生完整的類別組合為止。

## 六、實驗結果

本航空訂票查詢系統使用「音中仙」的 API(Application Program Interface)為連續語音辨識為辨識器，並結合中文文句翻語音系統 (TTS) 之 API 所發展而成。語音輸入後，辨識器輸出音節(syllable)的候選音，透過自然語言的處理，配出 n 個候選句子，再把這些句子輸入 HMM 語意模型中，把一些不合理的句子刪除，再求出其語意最高分數的句子，其所代表的語意便是我們所要。

在測試系統語音辨識部份時，我們先針對每個詞段類別的關鍵詞任意挑選唸 20 遍，再統計每個類別的正確率，結果如表一所示。

表一、單一關鍵詞的辨識率

類別	疑問詞	問候語	詢問詞	目標詞	From	To	肯定詞
正確率	85	90	95	95	80	75	85
類別	否定詞	時間前約 略語	時間後約 略語	動作詞	時間	班次	地點
正確率	85	90	90	90	85	80	100
類別	日期	人數	Filler	複誦	數字	人名	結束語
正確率	75	85	/////	/////	85	/////	95

在測試意圖 HMM 的效能方面，我們以人工蒐集錄音之航空查詢及訂位之對話語句共 82 個對話，其中包含一千多個對話語句，而後我們以人工把蒐集的語料，做人工的標示後，分別訓練四個意圖 HMM。而後我們從訓練語料中任意挑選 284 句作測試，測試結果如表二所示。而後，我們再由一些不包含在訓練對話語句的對話中，取出 172 句作人工標示後測試，其結果如表三所示。

表二、意圖 HMM 的效能測試表(close test)

	Booking Intention (110 句)	Query Intention (54 句)	Greeting Intention (33 句)	Ending Intention (42 句)	Other Intention (45 句)
Booking 個數	103	4	1	0	3
Query 個數	6	50	0	0	5
Greeting 個數	0	0	30	0	0
Ending 個數	1	0	2	42	0
Other 個數	0	0	0	0	37

表三、意圖 HMM 的效能測試表(open test)

	Booking Intention (32 句)	Query Intention (25 句)	Greeting Intention (23 句)	Ending Intention (57 句)	Other Intention (25 句)
Booking 個數	29	4	3	5	3
Query 個數	3	21	0	0	2
Greeting 個數	0	0	19	0	0
Ending 個數	0	0	1	52	0
Other 個數	0	0	0	0	20

表格的橫列，代表所用的測試語料的句數，縱列代表系統判斷屬於該種意圖的個數。其中 other intention 代表著不屬於前面四種意圖的句子，也就是被前面四種意圖所拒絕的句子，我們把他歸納為此種意圖。在所有辨識結果中，close test 平均約有 92%的辨識率，open test 約有 82%的辨識率，其中以結束意圖最為正確，原因在於結束用語自成一格，不太會與其他意圖的關鍵詞混淆所致。而其他意圖辨識最差，原因在於其他意圖中包含複誦部分，而複誦句子均是一些數字、英文字母等，且多夾雜很多非關鍵詞，導致辨識效果比較差。

在測試系統對於使用者的語句處理能力方面，我們從收集到的語料庫中找出 20 個



完整的對話做為測試，我們將整個對話根據語意的類別分成五個部份即問候、訂票與查詢(包括日期、時間、起迄點)、電話號碼、複誦及結束語分別做測試，且每個部份再分為 A、B 部兩部分，A 表輸入一次系統便可作出正確的回應，B 表輸入三次以內可作出正確的回應，結果如表四。

表四、回應能力統計表

		A		B	
		正確語句	正確率	正確語句	正確率
問候句	20	16	80%	18	90%
訂票/查詢	20	15	75%	17	85%
電話號碼	20	9	45%	10	50%
複誦	20	13	65%	14	70%
結束句	20	16	80%	18	90%

## 七、結論與討論

語意是一句話最重要的關鍵所在，所以判斷語意對於理解是最重要的部分。本論文在語意模型方面，提出隱藏式馬可夫模組(HMM)的觀念來建立與判別意圖的依據，不但結合了詞與詞之間的連接架構(N-Gram)，也考量到了合法句子的結構。

實驗的結果顯示，在測試的第一部份，針對每個類別的關鍵詞的正確率，其中關鍵詞長度過短或過長都會影響辨識結果，再者有關數字的關鍵詞辨識率較不高；其他的關鍵詞辨識率還不錯，皆可達到八、九成的效果。第二部分針對句子意圖的分析，意圖的正確與否，關係著對話系統的正確率，本論文所建構的意圖 HMM，在辨識率有 82%，對於對話系統處理有顯著的影響。第三部份針對整句的輸入做分析，我們分成問候句、訂票/查詢、電話號碼、複誦、結束句五個部分分別做測試，其中問候句、結束句的正確率較高；電話號碼與複誦這兩部份因有數字及英文字的辨識，若其中一個數字或英文字辨錯便算錯所以辨識率較低。整體上面，HMM 所構成的意圖模組，確實可以運用於意圖的辨識。

在未來的發展，人名的部份仍是一個需要研究與處理的部分。觀察語料的人名部份顧客與航空訂票人員的互動時，若是很普遍的表示方式，訂票人員很快能了解顧客所要表達的字，但通常這樣的來回並不是非常的順利，若是要使電腦能辨認與了解一般的姓名還要再研究。目前我們對姓名這部份尚未加以處理，這將是我們以後的研究重點。

致謝：

本研究部分由工研院電通所支持才得以完成，特此感謝。

#### 重要文獻(References)

- [1] Victor W. Zue, "Toward System that Understand Spoken Language," ARPA Strategic Computing Initiative, 1994.
- [2] Helen M., Senis B, and Victor Zue, et al. "WHEELS: A Conversational System in the Automobile Classification Domain," ICSLP '96 Vol. 1.
- [3] S. Bennacef and L. Lamel et al., "Dialog in the RAILTEL Telephone-Based System," ICSLP'96 Vol. 1.
- [4] Frank Seide and Andreas Kellner, "Toward an Automated Directory Information System," EuroSpeech'97 Vol. 3. pp.1327-1330
- [5] 李琳山, "國語對話技術之初步研究", 交談系統暨語境分析研討會, 中央研究院 1997
- [6] Chun-Jen Lee, Eng-Fong Huang, and Jung-Kuei Chen, "A Multi-Keyword Spotter for the Application of the TL Phone Directory Assistant Service," Proceedings of 1997 Workshop on Distributed System Technologies & Applications, pp. 197-202
- [7] Tung-Hui Chiang, Chung-Ming Peng, Yi-Chung Lin, Huei Ming Wang and Shih-Chieh Chien, "The Desigh of A Mandarin Chinese Spoken Dialogue System," in Proceedings of COTEC'98 , 台北 1998, pp. E2-5.1~E2-5.7

- [8] 張哲賓, 王駿發, “以音中仙為基礎之火車時刻表查詢系統”, 國立成功大學資訊工程研究所, 研究報告, 1996.
- [9] 劉定儀, “中文電話自動回應系統之研究”, 國立成功大學, 電機工程研究所, 碩士論文, 1993.
- [10] Hsien-Chang Wang, Jhing-Fa Wang, and Yi-Nan Liu, “A Conversational Agent for Food ordering Dialog Based on Venus Dictate,” Proceedings of ROCLING X International Conference 1997, pp. 325-334