

國立交通大學

資訊學院 資訊學程

碩士論文

詞風與情境判斷專家系統

Building a Ci Style and Scenario Diagnosis Expert System



研究生：許嘉妮

指導教授：曾憲雄 教授

中華民國九十六年六月

詞風與情境判斷專家系統

Building a Ci Style and Scenario Diagnosis Expert System

研究生：許嘉妮

Student: Chi-Ni Hsu

指導教授：曾憲雄

Advisor: Shian-Shyong Tseng



A Thesis

Submitted to College of Computer Science
National Chiao Tung University
in partial Fulfillment of the Requirements
for the Degree of
Master of Science
in
Computer Science
June 2007

Hsinchu, Taiwan, Republic of China

中華民國九十六年六月

詞風與情境判斷專家系統

研究生：許嘉妮

指導教授：曾憲雄 博士

國立交通大學 資訊學院 資訊學程碩士班

中文摘要

宋詞為我國傳統的四大韻文之一，其風格與情境抽象，往往必須透過文學專家的分析，方得以理解作品中的情感。本研究首先擷取詩詞專家知識，用以建立分析詞風與情境之知識，並建置成詞風與情境判斷專家系統 (Expert System)。

在本論文中，我們先建置分析宋詞韻文所使用的「宋詞概念階層」，以解決古文詞彙與現代漢語文出入的問題。然後擷取專家對於判斷宋詞詞風與情境的知識，分三個階段建構系統。第一階段為宋詞斷詞：【宋詞斷詞器】將宋詞按照節奏、典故、領字、構詞、對仗、專有名詞六大模組作精確斷詞，並由「斷詞模組規則知識庫」決定模組順序。第二階段是詞風判斷：透過詞彙所代表的語意，擷取專家知識以建置「詞風概念階層」，並依此設計規則集（形式數量、內容特徵、隱含特徵、詞調特徵），建構「詞風判斷規則知識庫」。第三階段為情境判斷：利用五種感官特徵，建置「情境概念階層」，其中包含三個階層（五感識別、感官概念、主體概念），並且建置為「情境規則知識庫」，以此判斷詞作的情境意涵。

經過實驗證明，在風格判斷上我們有80%左右的正確率，而在情境判斷的正確率上也有超過70%。在未來的工作，我們會持續增加並修訂所各種知識庫，使之能夠有更好的成果。

關鍵字：詞風、豪放、婉約、中文斷詞、宋詞

Building a Ci style scenario diagnosis expert system

Student: Chia-Ni Hsu

Advisor: Dr. Shian-Shyong Tseng

Degree Program of Computer Science

National Chiao Tung University

Abstract

The SongCi is one of Chinese traditional four major verses and the ancient usually expresses one's emotion with the wonderful words. But the style and scenario of SongCi are usually implicitly described and very difficult to understand. Therefore, to understand the emotion implied in the SongCi always needs the help of literature expert. In our research, we acquire the knowledge of Ci style and scenario from domain experts, and use it as the knowledge base of Expert System.

In this thesis, we constructed the SongCi Concept Hierarchy for Ci analysis firstly, because the term used in poem is different from modern Chinese article. Then, three phases are proposed to construct Ci Style and Scenario diagnosis Expert System based on the domain expertise. In first phase, SongCi Parser: uses the six modules to extract the nouns in each SongCi. In addition, we use the Parse Module Rule Knowledge Base to rank the order of the application of these modules. The second

phase is to construct Ci style diagnosis rule base: Form and Amount Rule Set, Content Feature Rule Set, Imply Feature Rule Set, and Rhythm Feature Rule Set. We construct different rule sets from concepts of terms and the domain expertise. According to these Rule Sets, Ci Style diagnosis Rules Knowledge Base can be constructed. The Third phase is Ci scenario diagnosis. We use the features of five human senses (Sight, Hearing, Smell, Taste and Dermal) to construct Ci Scenario Diagnosis Rules Knowledge Base. Three steps (Sense Distinguishing, Sense Word Concept and Object Concept) are proposed to find out the Ci scenario.

The experimental result shows that the accuracy is 80% at Ci Style Diagnosis and the accuracy is more than 70% at Ci Scenario Diagnosis. In the near future, we will continue to improve proposed knowledge base to increase the accuracy and the scalability of our system.

Keywords: Ci Style, Bold and Unconstrained(豪放), Graceful and Restrained (婉

約), Chinese Segmentation, SongCi

誌謝

經過兩年的時間，關於專家系統以及宋詞的研究，總算是暫時的告一個段落，並且促成這篇論文的完成。但非一己之力就能完成，而是眾人努力的成果，從早期的方向摸索、題目確定，到後期的系統架設及論文完稿，總是有師長及諸位夥伴的幫忙，為我指引方向及提供寶貴的意見，讓我學習到許多，也才能論文如期完成。

首先，感謝指導教授曾憲雄老師，認真和熱忱，是我從老師身邊體認到最深刻的特質，總是能夠耐心聆聽我荒謬的想法和意見，並且將其導引到正確的方向，在老師的指導之下著實讓我獲益良多。另外也要感謝口試委員洪宗貝教授、楊錦潭教授、曾秋蓉教授對於論文的寶貴建議，以讓論文更加趨於完善。

其次要感謝楊哲青學長，即使是在新婚期間，仍犧牲了個人時間，共同參與論文的討論、寫作及校稿。同時也要謝謝元昕、瑞鋒、俊銘、威勳、伊凡學長給予的建議與指導。還有一起努力的曉涵、芙明、昂叡、信男、雨杰、東權、智凱、敬斌、揚棋、裕傑大家互相打氣加油，以及負責幫忙各種相關的事宜。最後謝謝一直陪伴在身邊的家人與朋友們。

許嘉妮

2006 年 6 月

目錄

中文摘要	i
Abstract.....	ii
誌謝	iv
目錄	v
表目錄	vii
圖目錄	viii
演算法目錄	ix
第一章 緒論	1
第二章 研究背景與相關知識	4
2.1 宋詞介紹	4
2.2 宋詞詞風	5
2.3 宋詞情境	6
2.4 宋詞斷詞器	7
2.5 本體論(ONTOLOGY)	8
2.6 相關研究	9
第三章 詞風與情境判斷專家系統	11
3.1 系統架構	11
3.2 宋詞概念階層	12
3.3 詞風概念階層	14
3.4 情境概念階層	19
3.5 知識庫建置	23
3.5.1 斷詞模組規則庫	24
3.5.2 解歧義	25
3.5.3 詞風判斷規則庫建置	27
3.5.4 情境規則知識庫	30
第四章 系統實作與實驗	36
4.1 評估指標	36
4.2 系統建置與實驗設計	37

4.3 詞風判斷實驗.....	40
4.3.1 詞彙語意判斷.....	40
4.3.2 判斷語意判斷與節奏判斷.....	40
4.4 情境判斷實驗.....	41
4.4.1 情境判斷: 單就規則, 單一最佳解實驗	41
4.4.2 情境判斷:單就規則, 多選解實驗	41
4.4.3 情境判斷:加權值(Weight)實驗	42
4.4.4 分析比較.....	42
4.5 詞風與情境判斷實驗.....	43
第五章 結論與展望	45
參考文獻	47



表目錄

表 1 婉約與豪放詞風特徵.....	6
表 2 婉約與豪放下之情境.....	7
表 3 同義詞詞林大類.....	9
表 4 形式數量概念延伸.....	14
表 5 形式數量概念.....	15
表 6 內容特徵概念.....	15
表 7 視覺概念.....	20
表 8 聽覺概念.....	21
表 9 嗅覺概念.....	21
表 10 味覺概念.....	22
表 11 膚覺概念.....	22
表 12 “醉花陰”情境判斷結果 – 風格未知，No Weight.....	32
表 13 “醉花陰”情境判斷結果 – 風格未知，Weighted.....	33
表 14 “醉花陰”情境判斷結果 – 風格已知，No Weight.....	33
表 15 “醉花陰”情境判斷結果 – 風格已知，Weighted.....	33
表 16 詞風判斷結果.....	40
表 17 詞風判斷結果.....	40
表 18 情境判斷結果(單就規則-唯一最佳解).....	41
表 19 情境判斷結果(單就規則-多數候選解答).....	42
表 20 情境判斷結果(加入Weight).....	42
表 21 詞風與情境判斷結果(單就規則).....	43
表 22 詞風與情境判斷結果(加入weight).....	43

圖目錄

圖 1 詞風與情境判斷專家系統.....	11
圖 2 系統架構.....	12
圖 3 詞風概念階層.....	19
圖 4 情境分類.....	20
圖 5 情境概念階層.....	23
圖 6 規則知識庫.....	23
圖 7 解歧義.....	26
圖 8 「宋詞概念階層」語意距離.....	27
圖 9 詞風知識庫.....	29
圖 10 詞風判斷規則.....	29
圖 11 詞風與情境判斷專家系統架構圖.....	37
圖 12 系統首頁.....	38
圖 13 輸入斷詞結果.....	38
圖 14 詞風判斷.....	39
圖 15 情境判斷.....	39



演算法目錄

演算法 1: 隱含特徵演算法17



第一章 緒論

宋詞在中國古典文學上佔有極為重要的位置。著名詩詞評論家葉嘉瑩說：「詞是一種有很豐富想像餘地的作品，可以給讀者更自由的聯想。而詩則是言志的，即使寫得再好，可是卻都說出來了。」由此可知詞的優美在於其字裡行間隱含的意境以及說不出來的情感。宋詞與唐詩一樣必須遵守嚴謹的格律要求，古人必須利用簡短的文字表達無限的情感，目前針對宋詞格律檢測方面已發展許多系統，例如：羅鳳珠的【倚聲填詞系統】[22][23]。但詩詞更重要的價值在於作者所要表達的情感與意境。文學作品的詞風如人，是多采多姿的，即使是同一作者，由於不同時期與不同境遇，也會有不同詞風的作品，如：李清照南渡前後的不同，蘇軾、辛棄疾既有氣魄宏大，感情激昂的作品，也有纏綿悱惻、低迴婉轉的作品。在宋詞詞風分類上也見仁見智，但是，最基本分類為婉約和豪放。明朝張綏（《詩餘圖譜·凡例》）提出：「詞體大略有二：一體婉約，一體豪放。婉約者欲其詞調蘊藉，豪放者欲其氣象恢宏。然亦存乎其人。如秦少游之作，多是婉約，蘇子瞻之作，多是豪放。大約詞體以婉約為正。」此種分法，並不根據時代、地域或作者，純粹以作品的意境作為分類準則，這種分法獲得了明代以來文學界的普遍認同。

詞風為一首詞代表的風格，而情境則是詞作其主題中所使用的意象特色及形式內涵。洪惠華在《花間集的主題與感覺》[17]一書中，借用前人研究「詩」的方法來研究「花間集」的詞作，主要將詞作分為九種不同情境主題，分別為：浪漫的情愛、女性的姿態、怨曠與傷逝、詠物、入世的政治態度、出世的理想抱負、邊塞與地方風物、遊仙。但因為「遊仙」此情境只屬於花間集特有的意象，並不普遍使用於其他的詞作品，所以本論文將只使用其中八種做為情境主題分析結果。

宋詞如同一般的文學作品，其詞體風格以及意境感情難以解釋與量化，往往必須透過文學專家分析才能了解其中涵義。電腦也不易像讀者一樣能夠透過閱讀

或朗誦的過程中，直接了解其中的涵義。另外，詩詞不同於散文或抒情文章的鬆散架構，而是有其固定的格律，使得現行一般以計算詞頻為主的自然語言處理工具難以有效利用。因此在本篇文章中，我們提出【詞風與情境判斷專家系統】，來教導與協助學習者判別詞風與情境。在我們設計理念當中，基於字構成詞彙、詞彙構成句、句構成篇章的基礎，其中詞彙為代表語意的最小單位，所以系統以詞彙作為處理的對象。我們知道，在一篇豪放詞的作品裡面，其用詞必定帶有相對多數的豪放用語，例如：岳飛《滿江紅》中的"壯志飢餐胡虜肉，笑談渴飲匈奴血"，其中的壯志、胡虜、笑談、匈奴等詞彙。同樣的，在一篇婉約詞的作品當中，亦必帶有相對多數的婉約用語，例如：歐陽修之《蝶戀花》的前三句"庭院深深深幾許，楊柳堆煙，廉幕無重數"，其中的庭院、深深、楊柳、廉幕等詞。因此，我們以分析詞彙帶給人類的感覺作為知識基礎因而建立詞風判斷知識庫，並且從不同角度考慮詞彙的風格判斷，包括有對於事物數量的感受「形式數量特徵規則集」、考慮描寫題材的「內容特徵規則集」、考慮節奏的輕快或緩慢的「詞調特徵資料庫」、以及透過機器學習(Machine Learning)來延伸專家知識的「隱含(imply)特徵規則集」，以各種不同角度分析詞風。

另外，不同的詞風下往往會反應不同的情境內容，例如婉約詞風下的情境大多為描寫兒女情愛或是描寫離別與相思情愁，例如「浪漫的情愛」、「女性的姿態」、「怨曠與傷逝」，又如歌詠對愛情嚮往及美好景物的「詠物」；而在豪放詞風下的情境則大多屬於對國家之情、征戰情況或是悲痛亡國的「入世的政治態度」，或是隱退後及時行樂的放縱之情、表現男子曠達形象的「出世的理想抱負」，以及描寫江南閒適之情、邊塞之寒的「邊塞與地方風物」，而豪放詞風下的「詠物」則是以詠馬為主。一首詩或一闕詞之中，語言的最小單位是語詞，藝術的最小單位是意象，情境為全篇的構思，詩詞作家以五官感覺來開啟我們的審美經驗。因此我們試著利用一首詞作所使用的「感覺意象」來推論詞作所屬的情境，讓讀者能夠透過對詞風與情境的認識，來對詩詞本身以及作者當時的心境有更深刻的了解與感受。

在本篇論文當中，我們根據上述的各種專家知識，首先架構出「風格概念階層」和「情境概念階層」，並且依此建立詞風與情境判斷規則知識庫(Rule Base knowledge)，然後再導入最後的判斷機制，此判斷機制利用推論引擎(Inference Engine) DRAMA(國立交通大學知識工程實驗室先前所發展出的推論引擎機制 [6][7][21])來推論出我們所設計的「詞風與情境判斷專家系統」所要判斷的結果。

本論文的組織架構如下，在本章節說明研究的動機和研究方法，第二章介紹相關的研究背景，第三章針對系統架構作詳細的介紹以及專家系統知識庫的建置，第四章為系統實作以及實驗成果，最後為本論文的結論與針對未來研究的建議。



第二章 研究背景與相關知識

本章節將針對宋詞的詞風與情境類別以及分類知識加以介紹與說明，並對目前有關詩詞風格意境探討相關研究做一介紹與優缺點分析。

2.1 宋詞介紹

詞乃我國傳統的四大韻文之一，起源於隋唐，先是流傳於民間，大多數由民間不知名的人氏所寫，所用的音樂，融合了古代的樂曲、民間的歌謠、以及外來的胡樂，為歌女、伎工傳唱於歌樓酒肆之間，後來經由文人的修改，以及唐宋時國家音樂機關如教坊、大晟府的審音製曲，詞體大盛。詞在初興時因先有曲調，然後才有詞文，故稱「曲子詞」。由於詞可配合音樂歌唱，所以又稱為「倚聲」、「樂府」。亦因為句式長短不一而被稱為「長短句」，早期的詞是由詩歌發展而成，故又稱為「詩餘」。詞可以分為小令、中調和長調三種。字數在五十八字以內稱為小調，五十九字至九十字的稱為中調，而九十一字以上的便是長調或稱慢詞。

詞的分段稱為「闕」或稱「片」。大部分的詞可分為兩段，即兩闕（片），第一段稱為「上闕」或「上片」，第二段稱為「下闕」或「下片」。片與片之間稱為「過片」或「過變」，有的上片與下片的格式完全相同，有的則由不同的片組成。

每首詞皆有詞牌，詞牌即是一首詞的旋律、曲譜，每一詞牌都是一獨立的曲子，詞牌規定了詞的總字數、分段、句數、每句字數、平仄與用韻，作詞必須按詞牌填寫，不同的詞牌格式不一樣，唱法也不一樣。而所謂「詞調」，指的是一闕詞的曲調，每一個詞牌都有固定的詞調，包括固定的字數、句數、平仄、用韻。這就是明人徐師曾《文體明辨》所說的：「詞調規律有：調有定格，字有定數，韻有定聲」。

2.2 宋詞詞風

豪放與婉約，是宋詞兩大詞風，代表這兩者中詞人各有的不同個性與特色。豪放詞的代表作者為蘇軾、辛棄疾，婉約詞則為秦觀、李清照。現代文學專家利用分析宋詞的內容、形式及表現手法作為分類模式。

以內容來說，婉約詞多屬兒女之情、離別情緒[18]，例如李清照之《如夢令·昨夜風疏雨驟》描寫相思之情，或如柳永之《雨霖鈴·寒蟬淒切》描寫的離別情緒。豪放詞則多談論愛國之情，如岳飛之《滿江紅·怒髮沖冠》描寫家國之恨，或如蘇軾之《念奴嬌·大江東去》描寫貶官的身世感慨[12,13]。

另一不同處為婉約詞常用“設色語”[18]，意思是說用顏色來當形容詞，以周邦彥之《玉樓春》的中間四句為例，每一句皆使用一種顏色(赤、黃、青、紅)，設色語的使用為婉約詞風特色之一。

當時相候赤欄橋，今日獨尋黃葉路。煙中列岫青無數，雁背斜陽紅欲暮。

- 周邦彥《玉樓春》

另外，作品中所描述事物的形式數量大小[18]，也會影響人的感官，一般而言，事物如具有數量少、形體小、色彩素、聲音柔、味道淡、重量輕、速度慢等形式特徵會讓人覺得柔和，美學上稱為“優美”，指小巧、細膩、柔和，這與婉約詞風給人的感覺是相同的；而具有數量多、形體大、色彩艷、聲音粗、味道濃、重量重、速度快等形式特徵的事物容易引起人的強烈意味，美學上稱為“壯美”，指巨大、粗獷、豪放的美[15]，這則是豪放詞風給人的感受。所以事物數量的大小或是感官的強烈輕柔，也會直接影響文章的詞風。

詞最初是伴曲而唱的，曲子都有一定的旋律、節奏。這些旋律、節奏的總和就是詞調。詞調也與詞風有著密切關係，以詞句觀察，雙式句(句中的字數為複數)較為悠揚緩慢，適合婉約格調；單式句(句中的字數為單數)則較為暢快跳動，適合豪放格調。以豪放詞人常用的詞牌《水調歌頭》為例：

明月幾時有，把酒問青天。不知天上宮闕，今夕是何年。

我欲乘風歸去，又恐瓊樓玉宇，高處不勝寒。起舞弄清影，何似在人間。

轉朱閣，低綺戶，照無眠。不應有恨，何事長向別時圓。

人有悲歡離合，月有陰晴圓缺，此事古難全。但願人長久，千里共嬋娟。

- 蘇軾《水調歌頭》

上片:五、五、六、五、六、六、五、五、五

下片:三、三、三、四、七、六、六、五、五、五

上片(上半闕詞)中共九句中有六句為單式句;下片(下半闕詞)中共十句中有七句為單式句。因為單式句佔大多數所以唸起來節拍跳動，適合豪放詞的的壯烈激昂。婉約詞則相反。表一為針對宋詞詞風其內容、形式及表現手法做一扼要歸納。

表 1 婉約與豪放詞風特徵

	婉 約	豪 放
1	兒女之情或離別情緒	忠壯之情或身世感慨
2	女子容貌、形態、心思、閨房	論史談玄，師友之誼，田園風物，遨遊情態，家國之恨
3	多設色語	少設色語
4	取近景或細致之景(小,少,窄)	取遠景或壯闊之景(大,多,廣)
5	數量少、形體小、色彩素、聲音柔、味道淡、重量輕、速度慢	數量多、形體大、色彩艷、聲音粗、味道濃、重量重、速度快
6	詞調句法雙式句較多	詞調句法單式句較多

2.3 宋詞情境

根據洪華穗的《花間集的主題與感覺》一書中，對於婉約與豪放下的情境區分為：

表 2 婉約與豪放下之情境

婉 約	豪 放
浪漫的情愛	入世的政治態度
女性的姿態	出世的理想抱負
怨曠與傷逝	邊塞與地方風物
離情與別恨	詠物
詠物	

婉約詞風中的「浪漫的情愛」、「女性的姿態」、「怨曠與傷逝」、「離情與別恨」情境是以詠男女情愛兒女之情為主，用辭充滿柔性的驕態。「詠物」的情境則有詠動物、植物等各種不同的姿態展現，因為用詞上也以婉約詞彙居多，所以屬於婉約詞風下之情境。而豪放詞風下的「入世的政治態度」情境，則是描寫求仕中舉的喜悅之情，懷鄉愛國或是感慨悲痛亡國。「出世的理想抱負」情境則是當作者在政治抱負無法施展時，可能放情自己於酒樂之中，自我放逐，不管世事，是一種閒適與即時行樂的表現。「邊塞與地方風物」情境受到唐代邊塞詩的影響以描述南國風土人情的可愛之處，或是流離各地時歌詠邊塞風光與相思之苦的社會寫實詞。

如果利用實詞或名詞作為分析宋詞情境往往只能看到表面的事物，但是無法了解作者對於這些事物所聯想的情緒，所以我們試著從不同的角度，用更接近人類情感的感官知覺對宋詞作情境判斷。宋詞情境的形成，須由多種感覺互相組合，才能更為豐富。人類的感覺可分為視、聽、嗅、味、膚五覺，各種情境皆有其給人不同的感覺，所以詩詞專家也常以此作為區分情境的方式。

2.4 宋詞斷詞器

中文有別於英文之類的拼音文字，詞彙與詞彙之間並沒有空白的明顯區分，因此針對中文文章的處理，斷詞器為最基本也最重要的工作之一，錯誤或不精確

的斷詞，後續的文本處理則不可能正確。在本篇論文中，我們採用由許薰尹所發展的宋詞斷詞器[19]，此斷詞器有別於一般中文斷詞器只針對語體文處理，此系統針對宋詞的節奏停頓，以及宋詞特有的領字等特色進行斷詞，建構出一個規則式(Rule-Based)的斷詞方法對宋詞進行斷詞。在此系統當中，提供六大斷詞模組，包含有：專有名詞模組、領字模組、典故模組、構詞模組、節奏斷詞模組、對仗模組。每個模組各有其專責及功能：

- 專有名詞模組：使專有名詞不致被切分為單字詞，以免造成語意的混淆或錯誤。
- 領字模組：可切分出宋詞專有的虛字，虛字不會和其它的字組合成一個詞。
- 典故模組：可保留典故資訊，不致因斷詞而喪失原意。
- 構詞模組：可以補詞庫之不足，以解決部份未知詞問題。
- 節奏斷詞模組：則根據音律、句式對詞句做切割，以萃取其中的詞彙。
- 對仗模組：以詞句常用對仗加強氣勢的特色，輔助切分三字詞。

此外，系統利用規則中內涵的知識，讓系統可自動決定斷詞模組的順序。就整體而言，這個斷詞器所獲得的召回率與精確度兩者最佳可達 90% 的滿意度，資料的效度最高也達 90%，表示這種以宋詞節奏做為切分詞彙的方法具有一定的成效。宋詞斷詞器是以節奏斷詞模組為中心，除節奏模組之外，以領字模組對斷詞的影響最為顯著，能夠切割出宋詞的領字，就能夠大幅提高精確度、召回率和效度，也大大提升只使用一般詞庫的斷詞結果。

2.5 本體論(Ontology)

本體論[2][4][8]主要探討存在的本身，即一切現實事物的基本特徵。近年來，人工智慧及資訊技術相關領域的學者也開始將本體論的理念用在知識表達上[1]。Gruber曾定義：Ontology 是一種對某一個概念的詳細描述，包括對於概念、關聯、實體的描述。並清楚的定義其所欲表達的概念，主要的目的可用於知識的分享與再利用[3]。即藉由本體論中的基本元素：概念及概念間的關連，作為描述真實世界的知識模型。針對此一趨勢，W3C組織也開始定義了許多Ontology

的相關語言，如RDF、DAML+OIL、OWL等。

SUMO(Suggested Upper Merged Ontology, 建議上層共用知識本體)[5]是結合WordNet架構所建置的本體上層架構，由中研院歷史語言所維護，參考其SUMO概念架構，結合不同領域的知識本體，提供跨領域的資訊檢索，並可衍生出其他特殊領域的知識本體。但SUMO把概念視為節點，概念分類極為詳細，概念與概念間不單是上下隸屬的階層關係，也可以互相參考引用，在詞與的分類上沒有明確的界定，部分詞語間無法區隔歸類，使得語意分類上不易處理使用。

《同義詞詞林》[20]，或簡稱《詞林》，由上海外語學院梅家駒等人所編輯的中文寫作工具書，其中收錄將近七萬個現代漢語詞彙，主要用來解決中文寫作時所面臨詞窮的問題。其編排方式除了將所收錄的詞彙依據同義性質加以歸類整理，並考慮詞類，共分 12 大類，94 個中類，1428 個小類，小類之下再劃分成 3925 個詞群，其標註類別的格式為 Aa01，其中 A 為大類，a 為中類，01 為小類，其大類如下表 3：

表 3 同義詞詞林大類

大類	說明	大類	說明
A	人	G	心理活動
B	物	H	活動
C	時間與空間	I	現象與狀態
D	抽象事物	J	關聯
E	特徵	K	助語
F	動作	L	敬語

《同義詞詞林》收錄以現代漢語常見的一般語詞為主，也有一些古語詞，所以可以作為宋詞詞彙庫的參考工具書。且《同義詞詞林》對詞義分類架構明確，從大類到小類以至於以下的詞群，概念階層清楚且完整，所以常被拿來作為中文資訊處理的依據或參考。另外，在台灣發行的版本，亦加入台灣地區所通用的注音符號索引，提供更方便的查詢。

2.6 相關研究

在詩詞作品中，雙字以上之詞彙在某些情況下每各單字即可能代表一個詞意，所以我們必須透過構詞規則找出詞彙的主體，才能適當的將詞彙歸類。蘇豐

文在漢語詩的本體知識與語意檢索中[23]，提出根據詩句的特性以及《同義詞詞林》的架構歸納出一些詩句組成的綱要(schema)，知識綱要的初始建立是先由專家給予系統一些基本規則，之後系統會隨著知識的增加，自動產生出較複雜的綱要。利用這些綱要可找出詞彙的主體讓詞彙可以正確歸納到正確的概念中。

易勇運用機器學習(machine learning)技術分析宋詞風格[9]，利用單字詞出現率，建立豪放或婉約風格分類模式，但宋詞中常出現虛字、領字，如果只考慮單字詞會影響分類的準確性或是造成雜訊(noise)，而且詩詞常引用典故或專有名詞，加上單字詞詞義往往也與組合後詞意不相同，所以若能考慮詞彙的使用，更能建立準確的分類。

王迺仁在唐詩之詩風探勘中[10][11]，提出唐詩名詞概念階層，將語義繁雜的字詞轉換為概念精簡的名詞類別。此研究中以唐詩中的名詞作為唐詩詩風探勘的依據，屬於描寫事物上的探討，但是宋詞的意境屬於情感上的探討，不同的詞人即使描寫相同的事物也因為不同表達方式或不同心境而給人不相同的詞風感受。這也就是為什麼詞風的分析不應只侷限於名詞部分，必須包括所有相關的詞彙。另外，其探勘的方式，是採用分析詩作中名詞所使用的類別分佈，將大量的詩作加以分群探勘，這樣的方式可能需要相當大量的訓練資料才足以探勘出可靠性高的分群依據，而且這樣的分群依據及分群結果是否切合一般宋詞分類所須的類別令人質疑。

李良炎利用《基於詞聯接的詩詞風格評價技術》[16]方法，將詩詞風格分為豪放與婉約兩種風格。首先，他利用七種不同層級(-3, -2, -1, 0, 1, 2, 3)的意味語意決定詞彙的風格，然後透過詞聯接關係和語意價值的計算，來決定詞彙所構成的句子的意味語意，同樣的，利用這種方式計算句子所構成的段落的意味語意，以此類推，最後逐漸推算出整個篇章所具有的意味語意，並且利用所算出的意味語意來決定這篇詩詞所屬的風格為何。但在這之前必須以人工方式建立《詩詞熟語料庫》，將每各詞彙建構一個語意架構，但因為詞彙繁多且有不斷有新詞的出現，很難做到一個完整的語料庫。另外，這種的方式只能提供一種二分法的識別，無法針對詞作給予更多樣化的分類。

第三章 詞風與情境判斷專家系統

本章節中，我們將針對所提出的詞風與情境判斷專家系統的規劃及知識庫的建置步驟做介紹與說明。使用者輸入欲了解的詞作後，透過此專家系統作詞風與情境分析，最後將分析的結果回饋給使用者，如圖 1 所示。我們將此系統建立為一網路學習平台架構，提供使用者操作。

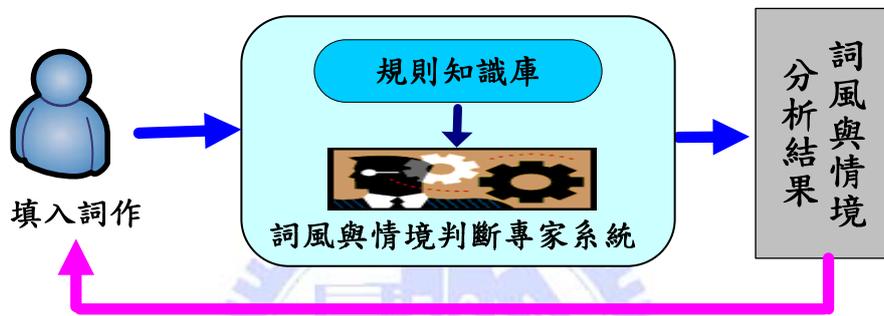


圖 1 詞風與情境判斷專家系統

3.1 系統架構

知識庫 (Knowledge Base) 的建置為專家系統的核心技術，本研究使用規則式知識(Rule-based Knowledge) 的方式來表示，以擷取詩詞專家判斷詩詞的方法來建置各項概念階層及規則知識庫。首先是輔助斷詞器的「斷詞模組規則知識庫」。接下來我們建置適用於詩詞分析使用的「宋詞概念階層」，並且根據詞風判斷的專家知識架構「詞風概念階層」，進而從當中擷取並建立判斷詞風的「詞風規則知識庫」。同樣的，我們根據情境判斷的專家知識架構「情境概念階層」，並且從當中建立判斷情境的「情境規則知識庫」，如圖 2 所示。我們將在後面章節詳細說明各部份的建置過程。

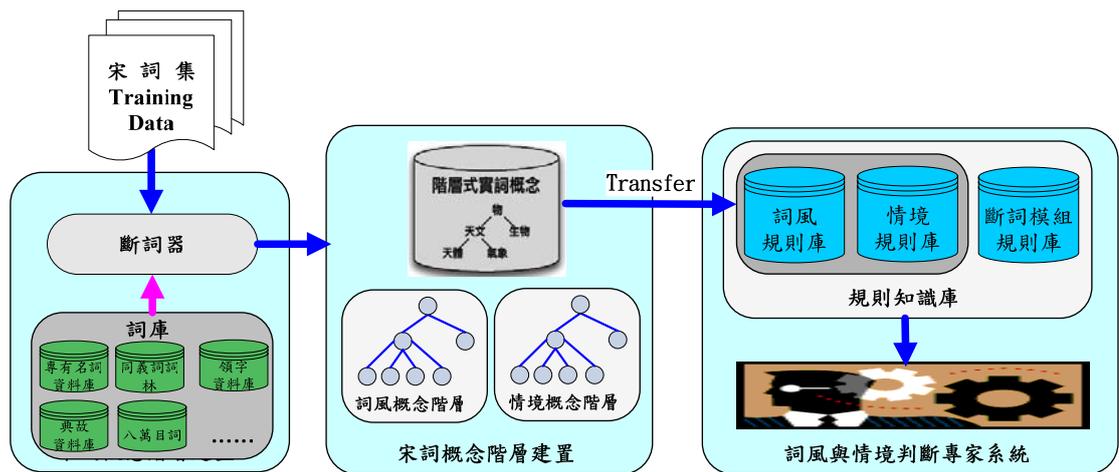


圖 2 系統架構

3.2 宋詞概念階層

詞風規則庫與情境規則庫的建置必須了解詞彙所屬的概念，然而《同義詞詞林》主要選收現代漢語詞，雖然也酌收了一些常見的方言詞與古語詞，不過仍有許多的古字詞仍未收錄，造成古文分析上的困難，但其明確且完整的分類架構，是值得學習參考引用的。所以本研究擴增《同義詞詞林》的古字詞數量，以及延伸其階層架構，使其成為更適用於詞風與情境分析時所使用的「宋詞概念階層」。

在擴增古字詞的過程中，由於詩詞所運用的多字詞彙在大多情況下每各單字即可能代表一個詞意(例如“紅花”所包含的“紅色”和“花”兩個詞意)，所以我們必須透過構詞規則找出詞彙的主體，才能適當的將詞彙歸類。本研究參考蘇豐文在「漢語詩的本體知識與語意檢索」[23]裡所提的構詞綱要，以及根據詩詞的特性以及《同義詞詞林》的架構歸納出一些詩詞詞彙組成的規則(Rule)，這些規則可幫助我們在建構宋詞概念階層時找出詞彙的主體，找出詞彙主體後則將此詞彙歸類於此主體的概念中。構詞規則的初始建立是先由專家給予系統一些基本規則，之後系統會隨著知識的增加，自動產生出較複雜的構詞規則要，並且再由詩詞領域專家(domain expert)校正。目前從斷詞的結果可歸納出一些基本的構詞規則，如下：

- (1) $E(\text{特徵}) + B(\text{物體}) = B(\text{物體})$ 、 $B(\text{物體}) + E(\text{特徵}) = E(\text{特徵})$ 。例：火紅
 $(B+E)=E$ 表示「火紅」的本質是「紅」， $紅火(E+B)=B$ 表示「紅火」的本質是「火」。
- (2) $E(\text{特徵}) + E(\text{特徵}) = E(\text{特徵})$ 。例：「細」與「長」分別都是特徵的形容詞，但當兩者結合在一起，「細長」也是一個形容特徵的詞彙。
- (3) $D(\text{抽象事物}) + D(\text{抽象事物}) = D(\text{抽象事物})$ 。例：萬(D)+年(D)=萬年(D)。
- (4) $B(\text{物體}) + B(\text{物體}) = B(\text{物體})$ 。例：星(B)+光(B)=星光(B)。
- (5) $B(\text{物體}) + D(\text{抽象事物}) = D(\text{抽象事物})$ 。例：[花草(B+B=B)+繽紛(D)]=D，
 意味著實體名詞加上抽象事物，可以合併成一個抽象事物的詞。
- (6) $B(\text{物體})/D(\text{抽象事物}) + C(\text{空間}) = C(\text{空間})$ ， $C(\text{空間}) + D(\text{抽象事物})/B(\text{物體}) = D(\text{抽象事物})/B(\text{物體})$ ；和(1)一樣，當 $B(\text{物體})/D(\text{抽象事物})$ 與 $C(\text{空間})$ 結合時，也往往合併成位於後者的概念，例：城東(B+C)=C 表示「城東」的本質是「東」， $東城(C+B)=B$ 表示「東城」的本質是「城」。

利用這些規則，將詞彙做完整的本體知識加註。雖然目前所擷取到的規則非常的稀少與簡略，只能得到粗略的構詞規則架構，雖然如此，有了粗略的規則，對詞的進一步語意分析已經提供了很多資訊與知識。

另外在延伸(extend)《同義詞詞林》的階層部分，因為詞風判斷實驗必須正確判斷出詞彙所屬的數量大小特徵，但在詞林的分類中只描述外型特徵，並不區分實際數量大小，所以這部份必須延伸詞林的類別以符合婉約或豪放的判斷。也就是說我們將《同義詞詞林》的分類架構依照原書分為大、中、小類三層架構，小類下再以同義原則延伸出第四層來劃分詞群。舉例來說“Ea03 大中小”為詞林中的小類，在此小類之下我們在將類別延伸成為“Ea03a 大”、“Ea03b 小”，也就是依照實際數量的大小或是多寡作為詞群分類依據。詳細概念以及部份詞彙可參考表四。

表 4 形式數量概念延伸

Ea03	Ea03a (小) 微 細 纖 細微 纖毫 微乎其微...	Ea03b (大) 巨 宏 碩 龐大 粗大 巨型...
Ea04	Ea04a (狹窄) 狹小 狹陋 侷促 狹隘...	Ea04b (廣闊, 寬敞) 寬闊 遼闊 廣大 浩蕩 無崖 連天...
Ea06	Ea06a (細) 尖細 纖纖 系若 削蔥 細挑...	Ea06b (粗) 肥大 粗壯 粗實 侏 裝 闊...
Eb01	Eb01a (少, 稀少) 零星 點滴 寥落 寥寥可數 秋毫...	Eb01b (多, 繁多) 浩大 博 大量 千千萬萬 如林...
Eb03	Eb03a (疏) 疏落 歷落 疏散 稀朗 稀稀拉拉...	Eb03b (密) 稠密 濃密 繁密 層層疊疊 密實...
Eb13	Eb13a (輕, 輕巧) 精巧 輕飄 清盈 翩然 翩翩 靈巧...	Eb13b (重, 笨重) 沈沈 壓秤 笨重 粗重 粗笨...
Eb21	Eb21a (近) 相近 在望 近在眉睫 一山之隔...	Eb21b (遠) 遙 遙遠 遙遙 迢遞 萬水千山...
Eb23	Eb23a (慢) 緩 徐 款款 遲緩 冉冉 悠悠...	Eb23b (快) 速 疾 快當 飛快 似箭 疾馳...
Ec11	Ec11a (清淡) 寡 清湯寡水...	Ec11b (醇厚, 油膩) 濃厚 醇 釀 濃郁...

3.3 詞風概念階層

擷取詩詞專家知識作為建構「詞風概念階層」的依據，以不同角度分析所屬的詞風，其中包含「型式數量」、「內容特徵」以及「隱含特徵」三部份，將分別敘述如下。

● 型式數量

根據事物的形式數量大小作為分類知識，建立形式數量的部份。首先標記「宋詞概念階層」中有關於數量描述的概念(concept)，目前共計九個概念，例如：Ea03(大小)、Ea06(粗細)等等，詳細概念如表 5 所示。

表 5 形式數量概念

婉 約	豪 放
Ea03a (小) 微 細 纖 細 微 纖 毫 微 乎 其 微..	Ea03b (大) 巨 宏 碩 龐 大 粗 大 巨 型..
Ea04a (狹窄) 狹 小 狹 陋 侷 促 狹 隘...	Ea04b (廣闊, 寬敞) 寬 闊 遼 闊 廣 大 浩 蕩 無 崖 連 天..
Ea06a (細) 尖 細 纖 纖 系 若 削 蔥 細 挑..	Ea06b (粗) 肥 大 粗 壯 粗 實 侷 裝 闊..
Eb01a (少, 稀少) 零 星 點 滴 寥 落 寥 寥 可 數 秋 毫..	Eb01b (多, 繁多) 浩 大 博 大 量 千 千 萬 萬 如 林..
Eb03a (疏) 疏 落 歷 落 疏 散 稀 朗 稀 稀 拉 拉..	Eb03b (密) 稠 密 濃 密 繁 密 層 層 疊 疊 密 實..
Eb13a (輕, 輕巧) 精 巧 輕 飄 清 盈 翩 然 翩 翩 靈 巧..	Eb13b (重, 笨重) 沈 沈 沈 壓 秤 笨 重 粗 重 粗 笨..
Eb21a (近) 相 近 在 望 近 在 眉 睫 一 山 之 隔..	Eb21b (遠) 遙 遙 遠 遙 遙 迢 遞 萬 水 千 山 ..
Eb23a (慢) 緩 徐 款 款 遲 緩 冉 冉 悠 悠..	Eb23b (快) 速 疾 快 當 飛 快 似 箭 疾 馳..
Ec11a (清淡) 寡 清 湯 寡 水...	Ec11b (醇厚, 油膩) 濃 厚 醇 釀 濃 郁...

● 內容特徵

根據所描寫事物內容不同的分類知識。豪放與婉約常描寫不同的情景與事物，所以透過詞彙所代表的不同內容，以及常使用於婉約詞風"設色語"的使用，建立內容特徵的部份。首先將「宋詞概念階層」中能代表詞風的概念作標記，例如:Ga02(憂愁,煩悶)、Ak01(英雄,硬漢,烈士)等等概念。另外，我們知道詩詞作品常使用專有名詞以及引用典故，這部份也屬於內容特徵規則的判斷。詳細概念以及部份詞彙可參考表 6。

表 6 內容特徵概念

婉 約	豪 放
Ah08(夫 妻 夫 妻)	Ae10(軍 官 將 士 軍 人 士 兵)

Bh02(花 花卉)	Af05(皇帝 后妃)
Bk(全身)	Af08(官吏)
Bp33(飾物 首飾 綉品)	Ag09(敗兵 俘虜)
Dc03(容貌 姿勢 步伐 裝束)	Ak01(英雄 硬漢 烈士)
Df04(感覺 感情)	Be01(陸地 原野 沙漠)
Df05(心意 心情 愛情)	Bo26(鞍轡 馬掌 鞭子)
Df06(苦心 憂心 怒火 怨氣)	Bo27(武器 槍 砲 刺刀)
Fc03(皺眉 撅嘴 接吻)	Bo28(彈藥 炸彈)
Ga02(憂愁 煩悶)	Bo29(弓箭 矛盾 劍)
Ga04(得意 惆悵)	De03(膽量 氣量 氣概)
Gb06(回憶 懷念 掛念)	Bo29(弓箭 矛盾 劍)
Ib21(開花 結果 凋謝)	Df11(志氣 意志 幹勁)
Ic04(嘆氣 發呆 沉默)	Di02(國家 行政區畫)
Ec01(紅 黃)	Ee06(忠誠 叛逆)
Ec02(藍 青 綠)	Ef01(太平 安定 動盪 混亂)
Ec03(紫 褐)	Fd09(行禮 下跪 磕頭)
Ec04(白 灰 黑)	Ga03(憤怒 息怒)
	Ha(政治活動)
	Hb(軍事活動)
	Hi06(送別 告別)
	Ic05(振奮 激昂)
	If(境遇)

● 隱含特徵

針對專家無法分類的詞彙，我們建立隱含特徵集來支援未知概念的判別。首先將專家已分類(豪放或婉約)的宋詞當作訓練資料(Training data)，將宋詞分類(豪放詞或婉約詞)與「宋詞概念階層」所對應的概念(Concept)紀錄為(CATEGORY_i, CONCEPT_j)，把已在"形式數量"以及"內容特徵"已定義過的

Concept 刪除，這些 Concept 是已經出現在專家的知識中，所以不再處理。另外，已經在"隱含特徵"定義過並且經過確認的，也不必再重複計算。系統所要處理的對象，是為前面三個部份皆未定義的概念。計算(CATEGORY_i, CONCEPT_j)的出現頻率，即某一 CATEGORY_i(豪放/婉約) 出現在哪些 CONCEPT_j，以出現的權重來將這些隱含的 CONCEPT_j 分類出來，做為「隱含特徵」，以延伸專家知識。隱含特徵演算法如下：

演算法 1: 隱含特徵演算法

Symbol Define:

Rule set A: Collection of CONCEPTs, 形式數量

Rule set B: Collection of CONCEPTs, 內容特徵

Rule set C: Collection of CONCEPTs, 隱含特徵

Database D: SongCi corpus

CONCEPT: 詞彙類別 Concept

Data Structure of CONCEPT_k{

CONCEPT_CODE //詞彙類別代號

CATEGORY //唐宋詞風(豪放或婉約)

TERM_i, i = 1 ~ j //此類別包含詞彙

}

Input: 宋詞 Cset, C_i 屬於 C, i=1~k;

每個 C_i 由詞彙陣列 Vector[TERM_x, CATEGORY] 組成

Output: Database C, Collection of CONCEPT_k

Algorithm:

Begin:

Step1: for each C_i in Cset {

for each TERM_x of C_i {

Step1.1: Map TERM_x to Database A, B, C's Concept,

Remove TERM_x from C_i if the TERM_x appears in Database A, B, C.

Step1.2: C_i' = Modified C_i, C_i' in the Cset'

//將 C_i 中出現在 DatabaseA, B, C 的詞彙移除，得到 C_i'

}

}

```

Step2: for each Ci' in Cset' {
    for each TERMx' of Ci' {
        Step2.1:if TERMx' is in Database X
            create new concept CONCEPTk' and add TERMx' to CONCEPTk',
            aggregate CONCEPTk' to Database C'
        //取 Ci' 中詞彙對應唐宋詞概念階層的 CONCEPTk'，得到 Database C'
    }
}

Step3: for each CONCEPTk' of Database C' {
    Step3.1:count number of TERMx' of CONCEPTk' belong to CATEGORY,
    set CONCEPTk's CATEGORY.
    //依 CONCEPTk' 中出現的 TERMx' 個數，來 assign CONCEPTk 的詞風類別
}

Step4: Merge Database C' to C.
    Recount the TERMx, reset CONCEPTk's CATEGORY.
    Reset the CONCEPTk of Database C.
    //Merge C' to C，重設 Database C 中的詞風類別
    return new Database C.

End

```

綜合以上的三個部份，我們架構出如圖 3 的「詞風概念階層」，並且根據此架構作為我們訂定詞風判斷規則集的依據。

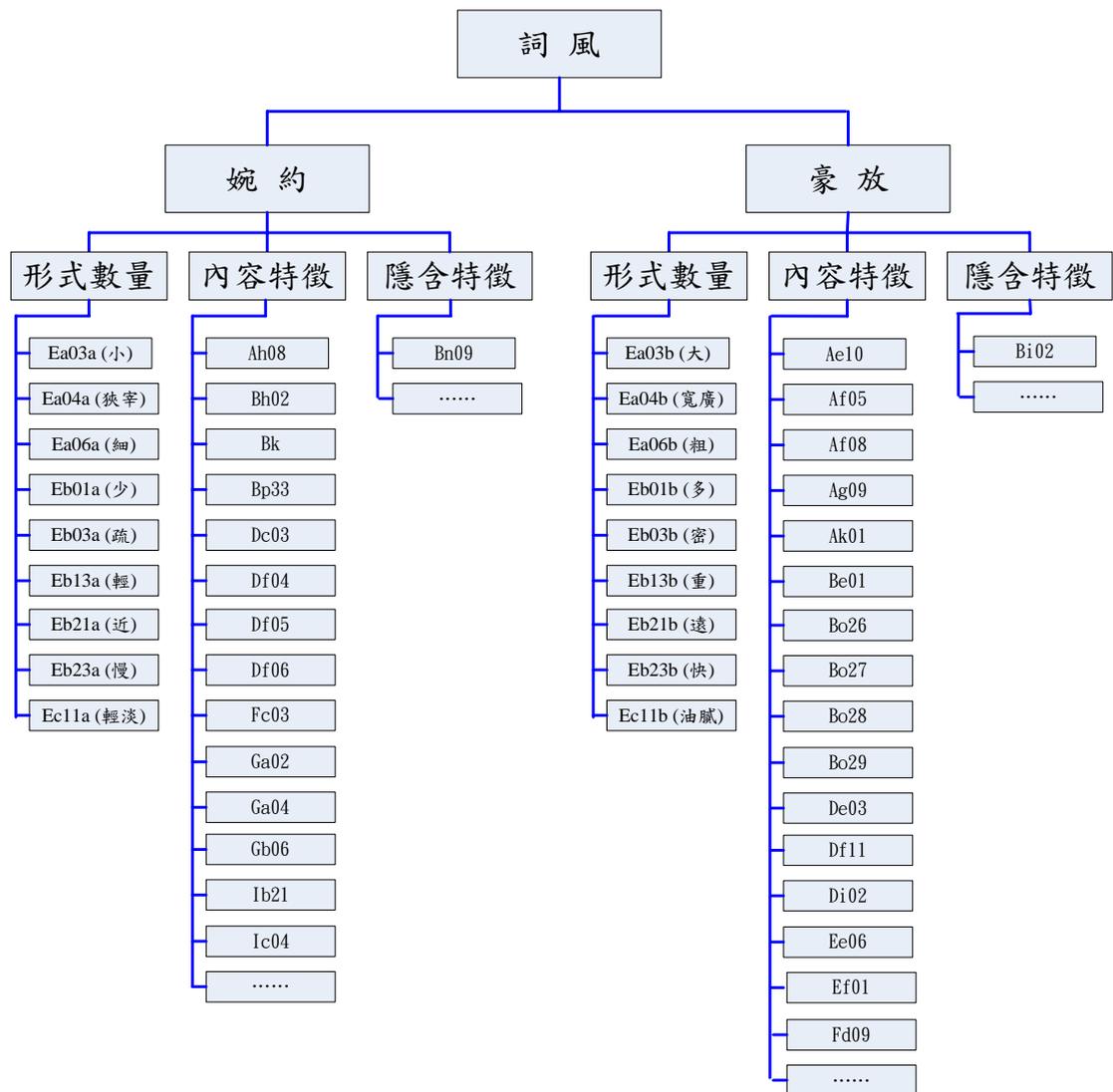


圖 3 詞風概念階層

3.4 情境概念階層

關於情境的概念階層，在我們的設計中，主要源自於洪華穗所著之《花間集的主題與感覺》一書，首先將八種不同的情境歸類在如圖 4 的詞風下，其中詠物情境在兩種詞風下皆有可能發生。

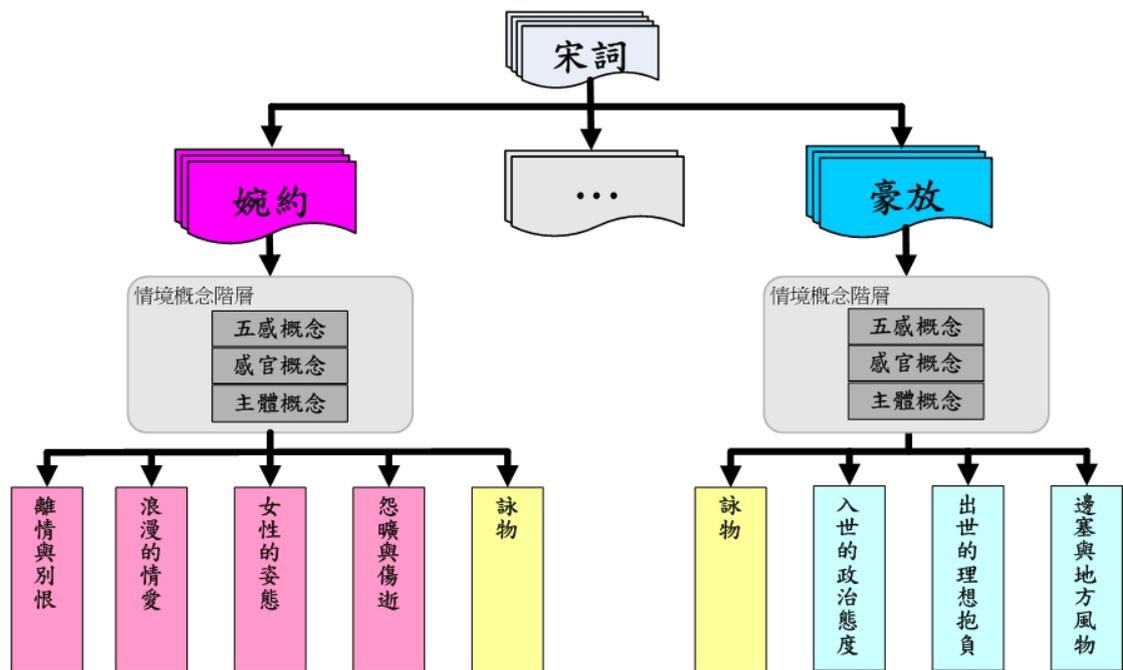


圖 4 情境分類

其次，我們參考文學專家的知識，根據五種感官(視覺、聽覺、嗅覺、味覺、膚覺)，以及每種感官下所區分的三個階層：**五感識別**(決定是否帶有感官意象以及屬於五感的哪一感)、**感官詞彙概念**(描述感官詞彙所屬概念及類別為何，例如視覺裡的紅色)、**主體概念**(識別詞彙的主體概念為何)，架構出「情境概念階層」，用以描述所有帶有五感的詞彙概念。我們針對每種感官意象，分別詳述如下：

● 視覺意象

在視覺的識別上，我們利用色彩做為視覺上的處理依據，視覺意象的詞彙包含形容顏色的詞彙或是字，然後進一步觀察此詞彙所包含的感官詞彙概念或類別為何，最後描述此詞彙的主體概念為何，如表 7 所示。

表 7 視覺概念

五感識別	感官詞彙概念	主體概念
Ec01、Ec02、Ec03、Ec04、Ec06	金、黃、紅、朱、白、玉、青、翠、碧、綠、藍、紫、彩...	妝裝飾、器物、酒杯、建物、居室、植物、動物、衣冠、首飾...

● 聽覺意象

在聽覺的識別上，約可分為動物、器物、人聲和自然景物四方面。對於動物、器物和自然景物，我們以這些事物本身作為判斷的依據，其中動物包含鳥類(Bi11~Bi13)、馬(Bi06)、昆蟲(Bi18, Bi19)，器物聲音則包括有屬於 Bp13(樂器)或是報時器的 Bp31(時鐘)概念，而自然景物的聲音則是以風、雨聲音居多，其所屬概念為 Bf01 及 Bf02，也可由詞彙所對應的概念作為依據。至於人聲部分則判斷是否有屬於概念 Dk11(話語)的詞彙。表 8 為綜合以上所得屬於聽覺感官的概念。

表 8 聽覺概念

五感識別	感官詞彙概念	主體概念
Bi、Bp13、Bp31、Bf01、Bf02、Dk11	動物	鳥類、昆蟲、馬...
	器物	樂器、報時器...
	自然景物	雨、風、水...
	人聲	話語、歌聲...

● 嗅覺意象

嗅覺主要以描述香味為主，因此識別嗅覺時，主要以香味的概念(Ec07)為主，另外再輔以馨、麝、檀、蕙等包含各種香味的字，當找出詞彙中含有「香」的概念後，再觀察所形容的主體為何，例如有:描寫座車(Bo21)的香車，描寫建築(Bn01、Bn02)相關的香閣、香殿，描寫灰塵(Bg08)的塵香，以及花香(Bh01~13)等等。表 9 為綜合以上所得屬於嗅覺感官的概念。

表 9 嗅覺概念

五感識別	感官詞彙概念	主體概念
Ec07、馨、麝、檀、蕙	香(Ec07)	人體、衣飾、妝、淚、器物、建物、自然、植物、檀香、爐香...
	其他(馨、麝、檀、蕙)	

● 味覺意象

味覺在人體五感當中，是感知範圍最狹窄的一個，僅限於人類的口舌，一般而言在現代用語多以酸甜苦辣等概念表示，然而在宋詞當中，這種情形並不多

見，反而以酒類相關最為常見，另外再輔以其他瓜果食物等概念，因此在味覺的描述依據，主要以酒的相關概念或字(酒(Br12)、杯、醉等)為主，另外再加上水果(Bh07)、五穀(Bh05)等概念、由於數量稀少且主要以主體本身為判斷依據，所以不再識別額外的主體概念，僅以酒類相關及食物相關作為其底層描述的依據。表 10 為綜合以上所得屬於味覺感官的概念。

表 10 味覺概念

五感識別	感官詞彙概念	主體概念
Br12、杯、醉、Bh05、 Bh07	Br12、杯、醉	
	水果(Bh07)、五穀(Bh05)...	

● 膚覺意象

膚覺詞彙主要是詞彙中帶有關於冷熱(Eb26)、乾溼(Eb27)、軟硬(Eb10)、輕重(Eb13)、纖細(Ea06)等的描述。因此我們可以很直接地去觀察詞彙裡是否包含這些概念作為判斷依據。其下一層規則的訂定，則以表示這些膚覺意象的字為主，例如冷熱(Eb26)概念裡的凍、寒、冷、涼、暖、熱等等。表 11 為綜合以上所得屬於膚覺感官的概念。

表 11 膚覺概念

五感識別	感官詞彙概念	主體概念
Eb26(冷熱)、Eb27(乾溼)、Eb10(軟硬)、Eb13(輕重)、Ea06(纖細)	凍、寒、冷、涼、暖、熱	妝、裝飾、器物、酒杯、建物、居室、植物、動物、衣冠、首飾...
	乾、溼	
	軟、硬	
	輕、重	
	薄、纖、細	

我們根據以上的觀念，架構出「情境概念階層」，利用三個階層來描述帶有五感意象的詞彙概念，如圖 5：

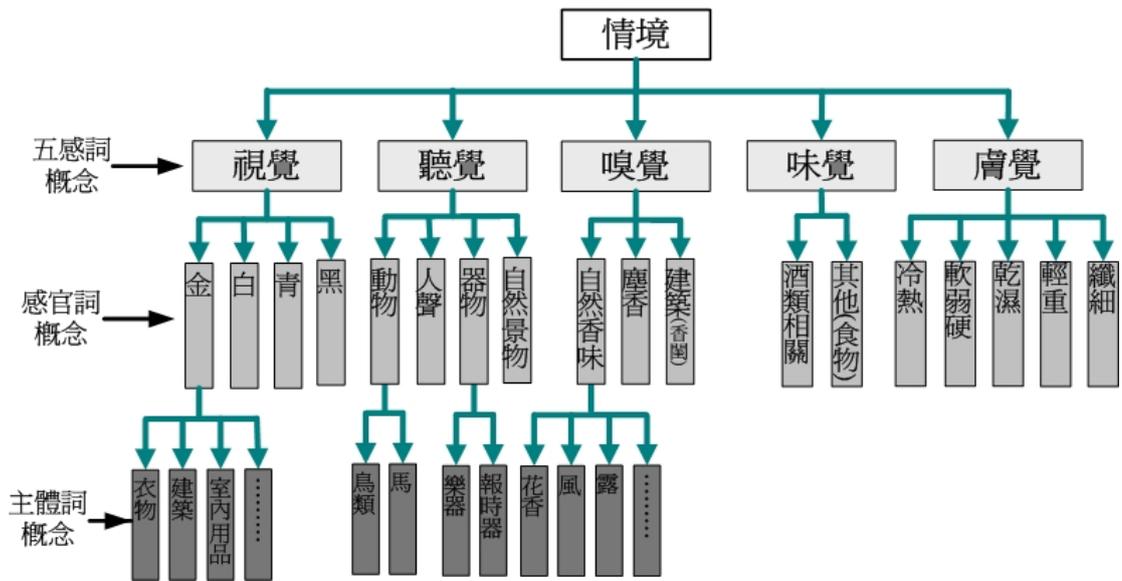


圖 5 情境概念階層

3.5 知識庫建置

規則式專家系統由兩大部分所組成，其一為規則式知識庫，另一部份則為規則推論引擎。推論引擎依據規則式知識庫裡之知識進行推論，所以規則式知識庫裡之建置則為規則式專家系統成敗之關鍵。本研究建立了「斷詞模組規則知識庫」、「詞風規則知識庫」、以及「情境規則知識庫」作為規則推論之依據。圖 6 為表示各知識庫間之關係，與其所參考的資料庫。

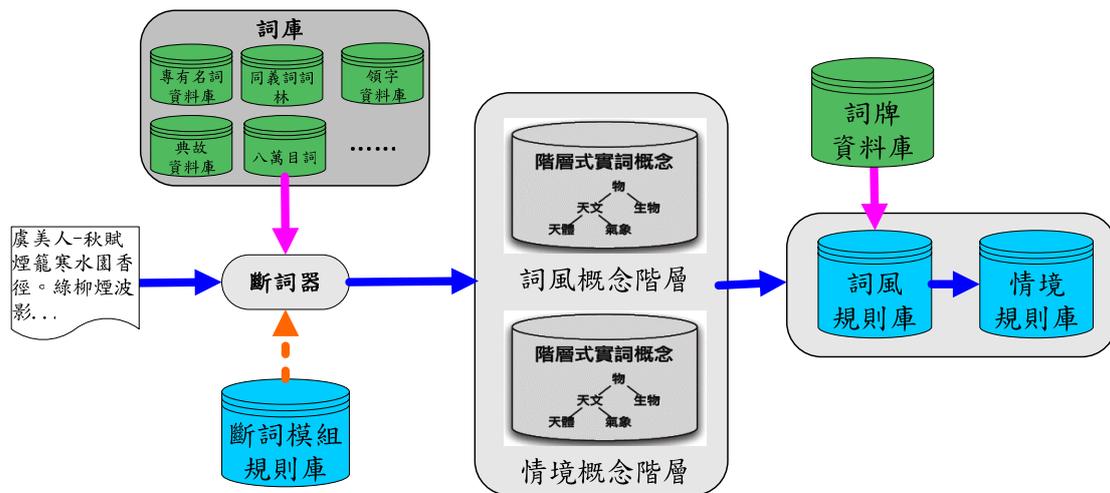


圖 6 規則知識庫

3.5.1 斷詞模組規則庫

宋詞斷詞器利用六大斷詞模組進行斷詞，這六個斷詞模組的順序若有變動，則會影響斷詞的精確度。「斷詞模組規則知識庫」是用來找出最佳的斷詞模組順序，目的是為了找出最多合法詞彙數量。首先系統會根據文學專家建議與觀察詞的句式，設定斷詞模組的優先權(priority)，如下列所示：

Priority (1): 專有名詞模組

Priority (2): 領字模組

Priority (3): 典故模組

Priority (4): 構詞模組

Priority (5): 節奏斷詞模組

Priority (6): 對仗模組

開始由優先權最高的模組先進行斷詞，如果正確斷詞的詞彙數量未達到所設定的門檻值(threshold)，則由次高優先權模組繼續斷詞，以此類推直到六大模組皆使用於斷詞。此時，正確斷詞的詞彙數量仍可能未達到所設定的(threshold)，這時則必須重新調整模組的優先權。調整的規則是以上次斷詞結果中，可以斷出最多正確詞彙的模組當作此次最高優先權的模組，以此類推重新定義模組的優先權，直到正確斷詞的詞彙數量達到所設定的門檻值(threshold)或無法排序出更好的順序後為止。透過此方式求得的斷詞結果，即為可斷出最多正確詞彙的模組順序。

斷詞模組規則

$C_i = \text{the Collection of Song } C_i \text{ Words}$

$UnParseC_i = \text{Remained words in } C_i \text{ after some parsers}$

$Term[] = \text{terms parsed from } C_i$

$Parser_i()$; Parser Modules, $i = 1\sim 6$

$Term_Pi[] = \text{Terms parsed by } Parser_i, i = 1\sim 6$

$SCCH(Term) = \text{Number of terms exist in 宋詞概念階層}$

$Sort_Parser()$; Sort Parsers according the number of each $Term_Pi[]$

IF $SCCH(Parser_1(C_i)) \geq P_Threshold$ THEN add $Term_P1[]$ into $Term[]$, return $Term[]$

```

IF SCCH(Parser_1(Ci)) < P_Threshold THEN add Term_P1[] into Term[] Parser_2(UnParseCi), add
TermP2[] into Term[]
IF SCCH(Term[]) ≥ P_Threshold THEN return Term[]
IF SCCH(Term[]) < P_Threshold THEN Parser_3(UnParseCi), add Term_P3[] into Term[]
IF SCCH(Term[]) ≥ P_Threshold THEN return Term[]
IF SCCH(Term[]) < P_Threshold THEN Parser_4(UnParseCi), add Term_P4[] into Term[]
IF SCCH(Term[]) ≥ P_Threshold THEN return Term[]
IF SCCH(Term[]) < P_Threshold THEN Parser_5(UnParseCi), add Term_P5[] into Term[]
IF SCCH(Term[]) ≥ P_Threshold THEN return Term[]
IF SCCH(Term[]) < P_Threshold THEN Parser_6(UnParseCi), add Term_P6[] into Term[]
IF SCCH(Term[]) ≥ P_Threshold THEN return Term[]
IF SCCH(Term[]) < P_Threshold THEN Sort_Parser()
IF no change after Sort_Parser() THEN return Term[]
IF changed after Sort_Parser() THEN reset Term[]&Term_Pi[]

```

3.5.2 解歧義

在詞風與情境判斷前，必須將斷詞結果的詞彙對應(mapping)到正確的概念，此情況下可能會有詞彙歧義(Ambiguity)的問題產生。所謂的詞彙歧義是指一詞多義，也就是說相同詞彙使用在不同詞句中，而有不同的解釋，當我們分析文學作品時，必須判別正的確解釋才能了解作品真正的涵義，而我們提出的解歧義方式有對仗以及計算上下文語意距離兩種方式。但某些情況下，對於沒有對仗資訊的詞彙，對仗資訊仍無法解決歧義問題，或者上下文語意距離無法提供正確詞意時，可透過詩詞專家的詞義校正與標註，以人工判別的方式解釋歧義詞彙。圖7為解詞彙歧義的過程以及所參考的資料庫。以下則針對所提出的解歧義方式做進一步說明。

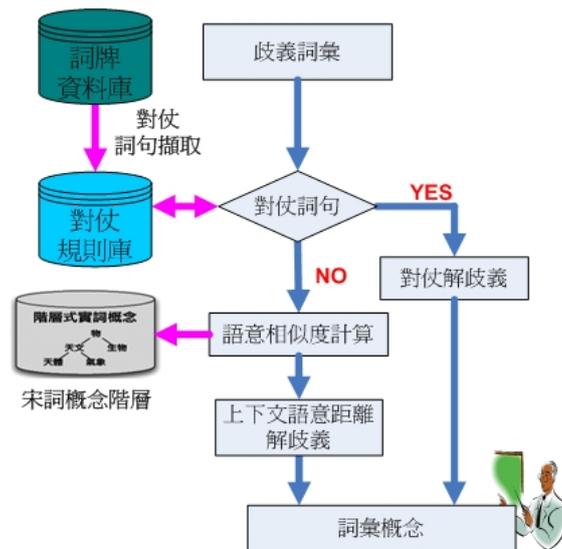


圖 7 解歧義

● 對仗

對仗是指兩個句子間的一種關係，這種關係必須滿足下述的要求：兩個句子的字數、句讀、語法結構、相同位置上的詞語的詞性，都要完全相同；而且兩句的平仄規律是“相對”的。滿足這種關係的兩個句子就構成一個對仗結構。對仗是中國古典詩詞的重要手段之一，但詞和詩不太相同，詞句並不要求一定要對仗，但詞的某一些句子中，只要相連兩句字數相同，經常使用對仗，利用此種特性可協助我們當作解歧義的方法。而對仗知識的擷取，則是透過詞牌資料庫取得，詞牌資料庫會記錄目前所有宋詞的詞牌，而詞牌中會標示必須對仗的句子以及詞彙。舉例來說《鷓鴣天》詞牌的上闕第三四句、下闕第一二句一般要求對仗，以李清照的詞為例：

秋/已盡，日/猶長。

李清照《鷓鴣天》

已知‘秋’與‘日’有對仗關係，而‘日’的解釋可能為天體概念的太陽，或代表時間與空間概念的日子，假設已知秋為秋天，秋天屬於時間與空間概念詞彙，所以我們可判斷這裡的‘日’為日子的意思。

● 上下文語意距離

詞語距離有兩類常見的計算方法，一種是根據本體論知識 (Ontology) 來計算，一種利用大規模的語料庫進行統計。根據本體論知識計算詞語語義距離的方法，一般是利用一部同義詞詞典 (Thesaurus)，一般同義詞詞典都是將所有的詞組織在一棵或幾棵樹狀的層次架構中。我們知道，在一棵樹形圖中，任何兩個結點之間有且只有一條路徑。於是，這條路徑的長度就可以作為這兩個概念的語義距離的一種度量。在解歧義方面，我們參考余曉峰等所發表的“一種基於《知網》的漢語詞語詞意消歧方法”[14]，所不同的是我們利用我們所建構的「宋詞概念階層」來計算漢語詞語之間的相似度，如圖 8 所示。簡單地說，針對某個歧義詞，我們首先找出其所包含的各種詞意概念在「宋詞概念階層」裡的位置，然後分別計算這些概念與這個歧義詞上下文最近的實詞的相似度(語意距離)，取其中相似度最大者(距離最短者)為其實際概念。例如：Aa01 與 Aa02 的語意距離=2，Aa01 與 Ab02 的語意距離=4

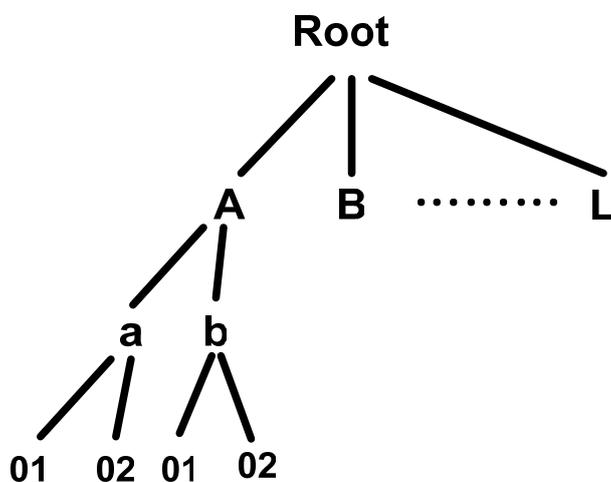


圖 8 「宋詞概念階層」語意距離

3.5.3 詞風判斷規則庫建置

針對我們設計的專家系統的詞風判斷部份，我們根據先前提到的「詞風概念階層」以及詞調特徵，設計(A)形式數量規則集、(B)內容特徵規則集、(C)隱含特徵規則集以及(D)詞調特徵規則集等四個不同角度的規則集(Rule Set)。

(A)形式數量規則集：根據在「詞風概念階層」裡的「形式數量」概念的特性，建立形式數量規則知識庫。在應用時擷取宋詞斷詞器之斷詞結果中出現關於數量描述的詞彙，並將此詞彙對應(mapping)到所屬概念，進而決定此詞彙所代表的詞風為何。

(B)內容特徵規則集：根據在「詞風概念階層」裡的「內容特徵」概念的特性，建立內容特徵規則知識庫。在應用時將詞題與詞作內容斷詞後的結果依序對應至概念上，檢查概念所代表的詞風即可知道此詞彙所代表的詞風。

(C)隱含特徵規則集：根據在「詞風概念階層」裡所建立的「隱含特徵」詞彙概念的特性，建立隱含特徵規則知識庫。和(A)(B)規則集相同，在應用時將斷詞後的結果依序對應至隱含特徵演算法所建立的概念上，檢查其概念所可能的詞風即可知道此詞彙所代表的詞風。

(D)詞調特徵規則集：詞調對詞風來說有定型的作用，可作為判斷詞風的依據。我們已知雙式句較為悠揚緩慢，適合婉約格調；單式句則較為暢快跳動，適合豪放格調；其中為雙式句或單式句則決定在於一個句子的字數，整首詞作若句子的字數以雙數較多則為雙式句，反之則為單式句。詞牌都有固定的詞調，包括固定的字數、句數、平仄、用韻，所以計算詞牌中每句的字數，則可知道此詞牌是屬於雙式句或單式句詞牌，進而可以當作判斷詞風的參考。

圖 9 為各種規則集在詞風知識庫裡所扮演的角色，其中形式數量規則集所處理的是用詞當中影響詞風的形式數量詞彙，而內容特徵規則集和隱含特徵規則集則是處理詞彙所屬的內容，詞調特徵規則集則依據宋詞專有的詞牌結構，判定可能的詞風。以詩詞專家對於詞風判斷的知識來分析，形式數量規則與內容特徵規則為重要分類依據，所以設定為第一階段，第二階段則加入透過系統學習後所得到的隱含特徵規則，當以上資訊仍無法判定詞風時則加入第三階段詞調特徵資料庫作判斷。假設:A=形式數量規則集的判斷結果、B=內容特徵規則集的判斷結果、C=隱含特徵規則集的判斷結果、D=詞調格式；而"-"代表婉約詞風，"+"代表豪放詞風，"0"代表中性詞風，中性詞風表示整首詞作中沒有出現能代表某一詞風的詞彙或是婉約詞風詞彙與豪放詞風詞彙數量相等時。而圖 10 則為各種規則

集的推演過程的架構圖。

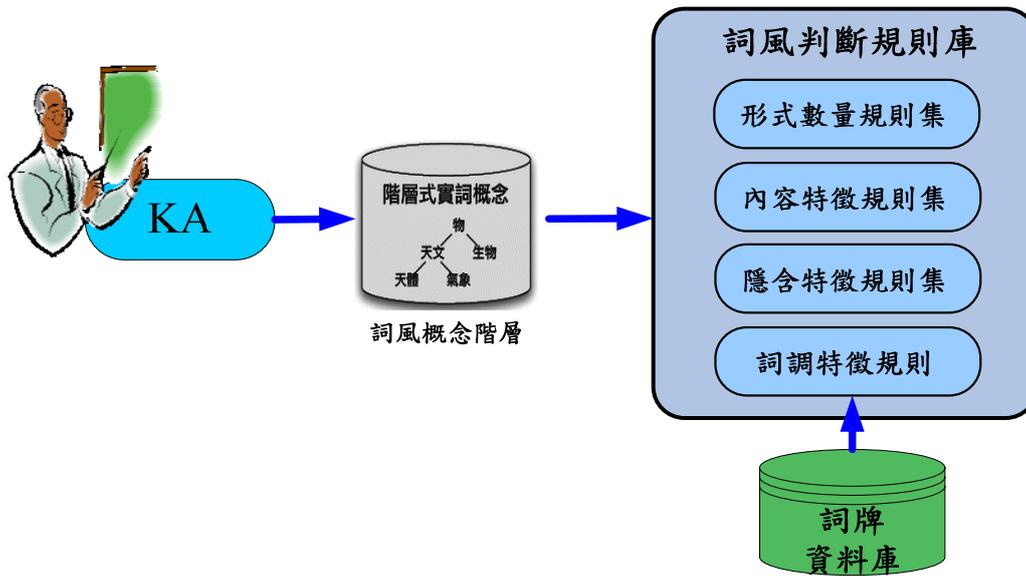


圖 9 詞風知識庫

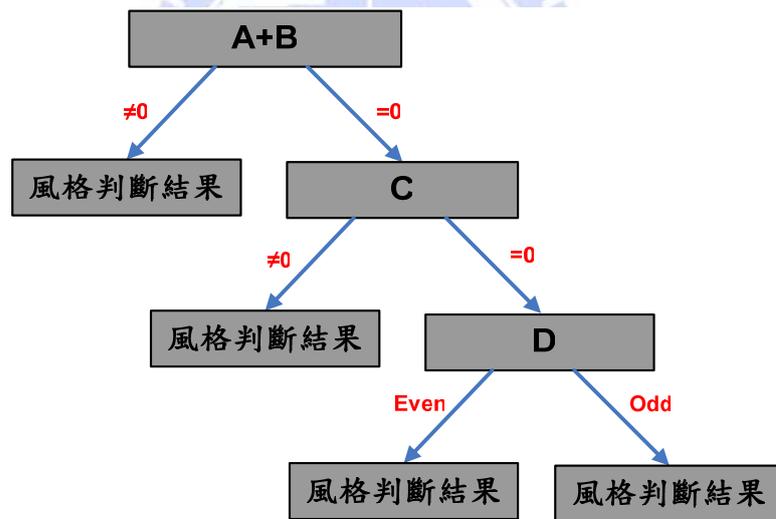


圖 10 詞風判斷規則

詞風判斷規則如下：

詞風判斷規則

```
//形式數量值(A)計算 (Rules in 型式數量規則集)
IF Concept_Term = Ea03a(小) THEN subtract 1 from A
IF Concept_Term = Ea03b(大) THEN add 1 into A
IF Concept_Term = Ea04a(狹窄) THEN subtract 1 from A
IF Concept_Term = Ea04b(廣闊, 寬敞) THEN add 1 into A
.....

//內容特徵值(B)計算 (Rules in 內容特徵規則集)
IF Concept_Term = Ah08(夫、妻、夫妻) THEN subtract 1 from B
```

```

IF Concept_Term = Bp33(飾物、首飾、綉品) THEN subtract 1 from B
.....
.....
IF Concept_Term = Ae10(軍官、將士、軍人、士兵) THEN add 1 into B
IF Concept_Term = Hb(軍事活動) THEN add 1 into B
.....
.....
//隱含特徵值(C)計算 (Rules in 隱含特徵規則集)
Count_C:
    IF Concept_Term == Bn09(陽台、院子) THEN subtract 1 from C
    .....
    IF Concept_Term == Bi02(獅、虎、豹、象) THEN add 1 into C
    .....

//詞調特徵值(D)計算 (Rules in 詞調特徵規則集)
Count_D:
    IF CiPei == “一剪梅” THEN D = even
    IF CiPei == “水調歌頭” THEN D = odd
    .....
    .....

// 詞風判斷規則
IF A+B < 0 THEN output “婉約”
IF A+B > 0 THEN output “豪放”
IF A+B == 0 THEN count_C, Rule_C
Rule_C:
    IF C < 0 THEN output “婉約”
    IF C > 0 THEN output “豪放”
    IF C = 0 THEN count_D, Rule_D
Rule_D:
    IF D == “even” THEN output “婉約”
    IF D == “odd” THEN output “豪放”

```

3.5.4 情境規則知識庫

針對我們設計的專家系統的情境判斷部份，我們根據先前提到的「情境概念階層」，設計出情境規則知識庫，對於斷詞後的詞彙，我們對應「情境概念階層」所架構的三個階層，分成三個階段判斷每各感官詞彙的情境意涵；第一階段：判斷此詞彙是否包含感官詞以及屬於感官意象的哪一類(ex.視聽嗅味膚)；第二階段：找出感官詞彙所屬概念及類別為何(ex.紅、綠、輕、重...)；第三階段：辨別詞彙描寫的主體概念為何(ex.人、動物、器物...)，以判斷規則是否成立，並且計算此規則分別提供各種情境多少的可能性得分。

舉例來說，當判斷“紅花”詞彙時，因為“紅花”一詞並不是顏色，所以必須將詞彙再進行切詞成為“紅”與“花”，已知“紅”為符合視覺中的顏色概念(concept)，而“花”則是符合“植物”概念，所以可判斷“紅花”符合視覺上的規則，並且是以紅色來形容植物，其識別規則如下：

```
IF concept is (Ec01 or Ec02 or Ec03 or Ec04 or Ec06) in Term THEN RuleSet_Sight  
(判斷感官字詞)
```

```
RuleSet_Sight:
```

```
IF Sense_Word == “紅” THEN Rule_Red (判斷顏色)
```

```
Rule_Red:
```

```
IF Object_Concept == “Bh(植物)”, THEN score to Scenario (主體字詞，得分)
```

另外必須注意的是，一個詞彙可能會兼具兩種以上的感官意象，例如“微風”兼具了聽覺的“風”以及膚覺的“微”，因此當某個詞彙在某種意象規則集判斷出符合的規則時，不能直接跳出去判斷下一個詞彙，必須繼續判斷是否屬於其他感官詞彙，直到五種規則集都判斷為止。

在一關詞當中所有的詞彙經過以上的計算，我們得到所有的感官詞針對各種情境提供多少的可能性得分，最後以最高可能性得分的情境則為此關詞所屬的情境。

此外，在情境規則知識庫當中，每一條規則在不同情境下是會有不一樣的重要性，所以我們從【花間集的主題與感覺】所統計的資料當中，(浪漫的情愛37首、女性的姿態45首、怨曠與傷逝104首、離情與別恨182首、詠物26首、入世的政治態度29首、出世的理想抱負26首、邊塞與地方風物34首-共483首)，利用各種感官詞在以上各種情境當中被使用的次數作為系統的訓練資料(training data)，並且拿來調整規則之權重值(weight)，讓每條規則在不同情境下會有不同的重要性，舉例來說，“聽覺+器物+樂器”使用在浪漫的情愛情境下共3次，女性的姿態為10次，怨曠與傷逝為9次，離情與別恨為20次，詠物為6次，另外我們必須考慮書中在每種情境裡所統計的樣本數量，分別為每種情境的數量乘上一固定常數，使樣本趨於平

衡，其中浪漫的情愛常數為5，女性的姿態為4，怨曠與傷逝為2，離情與別恨為1，詠物為7，所以依序每各情境下“聽覺+器物+樂器”使用的權重值為：

$$\text{情愛: } 3*5 / (15+40+18+20+42) = 15/135 = 0.11$$

$$\text{女性: } 10*4 / (15+40+18+20+42) = 40/135 = 0.30$$

$$\text{怨曠: } 9*2 / (15+40+18+20+42) = 18/135 = 0.13$$

$$\text{離別: } 20*1 / (15+40+18+20+42) = 20/135 = 0.15$$

$$\text{詠物: } 6*7 / (15+40+18+20+42) = 42/135 = 0.31$$

於是，我們便根據此數據，訂定出當含有“聽覺+器物+樂器”的詞彙出現時，其對於每個情境所提供的可能性得分，如下所示：

```

■ Rule_MusicalInstrument {
  ● IF Object_Concept == “Bp13(樂器)” THEN {
    Add 0.11 to Scenario_A;
    Add 0.30 to Scenario_B;
    Add 0.13 to Scenario_C;
    Add 0.15 to Scenario_D;
    Add 0.31 to Scenario_E;
  }

```

推論引擎依照制定完成的規則做推演，當詞彙概念符合所制定的規則時，則會開始計算此規則在每一情境下所給予的值，依照符合的程度統計每個情境的得分，最後將情境結果顯示給使用者。以下為一闕詞在各種不同的方法下，運算的過程與結果：

詞風未知：未經詞風判斷事先得知其詞風，其範例如表 12、13 所示。

表 12 “醉花陰”情境判斷結果 – 風格未知，No Weight

符合規則	情愛	女性	怨曠	離別	詠物	入世	出世	風物
視覺+金+器物	1	1	1	1	1	1	1	1
視覺+黃+自然	0	0	1	1	0	0	0	1
視覺+黃+植物	0	0	1	1	1	1	1	0
視覺+玉(白)+器物	0	1	1	1	1	0	1	0
聽覺+自然+風	1	1	1	1	1	1	1	1

嗅覺+香+自然	1	1	1	1	1	1	1	1
味覺+酒	1	1	1	1	1	1	1	1
膚覺+冷熱+涼	1	0	1	1	1	1	1	0
膚覺+纖細+薄	1	1	1	1	0	1	0	0
9	6	6	9	9	7	7	7	5

表 13 “醉花陰”情境判斷結果 – 風格未知，Weighted

符合規則	情愛	女性	怨曠	離別	詠物	入世	出世	風物
視覺+金+器物	0.26	0.08	0.09	0.08	0.09	0.15	0.22	0.04
視覺+黃+自然	0.00	0.00	0.24	0.12	0.00	0.00	0.00	0.65
視覺+黃+植物	0.00	0.00	0.18	0.07	0.51	0.11	0.13	0.00
視覺+玉(白)+器物	0.00	0.17	0.35	0.17	0.15	0.00	0.15	0.00
聽覺+自然+風	0.05	0.07	0.14	0.11	0.19	0.14	0.19	0.10
嗅覺+香+自然	0.06	0.06	0.12	0.16	0.13	0.16	0.16	0.15
味覺+酒	0.14	0.02	0.03	0.07	0.04	0.17	0.36	0.16
膚覺+冷熱+涼	0.14	0.00	0.22	0.08	0.19	0.17	0.19	0.00
膚覺+纖細+薄	0.09	0.07	0.33	0.28	0.00	0.22	0.00	0.00
9	0.74	0.48	1.70	1.15	1.31	1.13	1.40	1.09

詞風已知：事先經由詞風判斷得知其詞風為婉約，其範例如表 14、15 所示。

表 14 “醉花陰”情境判斷結果 – 風格已知，No Weight

符合規則	情愛	女性	怨曠	離別	詠物
視覺+金+器物	1	1	1	1	1
視覺+黃+自然	0	0	1	1	0
視覺+黃+植物	0	0	1	1	1
視覺+玉(白)+器物	0	1	1	1	1
聽覺+自然+風	1	1	1	1	1
嗅覺+香+自然	1	1	1	1	1
味覺+酒	1	1	1	1	1
膚覺+冷熱+涼	1	0	1	1	1
膚覺+纖細+薄	1	1	1	1	0
9	6	6	9	9	7

表 15 “醉花陰”情境判斷結果 – 風格已知，Weighted

符合規則	情愛	女性	怨曠	離別	詠物
------	----	----	----	----	----

視覺+金+器物	0.43	0.13	0.15	0.13	0.15
視覺+黃+自然	0.00	0.00	0.67	0.33	0.00
視覺+黃+植物	0.00	0.00	0.24	0.10	0.67
視覺+玉(白)+器物	0.00	0.21	0.41	0.21	0.18
聽覺+自然+風	0.09	0.13	0.24	0.20	0.33
嗅覺+香+自然	0.11	0.12	0.22	0.30	0.26
味覺+酒	0.45	0.07	0.11	0.24	0.13
膚覺+冷熱+涼	0.22	0.00	0.35	0.13	0.30
膚覺+纖細+薄	0.12	0.10	0.43	0.36	0.00
9	1.43	0.75	2.82	1.99	2.01

本系統以 Meta Rule 的方式架構，其優點為當往後判斷因素有所擴充或修訂時，可以直接針對目標規則修訂而不需調整整個架構。以本系統架構為例，我們將判斷的過程切分為三個階段，在第一階段判斷是否帶有感官詞並決定所屬感官，然後在第二階段決定感官詞概念，最後在第三階段判斷整個詞彙的本體概念並決定對各種情境所提供的可能性得分，因此當系統執行時，我們是用逐步推論的過程，去確定我們所判斷的詞彙為各種情境提供多少的可能性。另外，當我們需要加入新的規則時，只需視規則所牽涉的層級來調整即可，而不必將整個規則集。綜合以上的各種知識與想法，我們歸納整理出如同以下的規則集：

情境判斷規則

```
// Layer 1 Rule Set
  IF concept [Ec01|Ec02|Ec03|Ec04|Ec06] in Term THEN Rule_Sight
  IF concept [Fc10|Bi|Bp13|Bp31|Bf01|Bf02] in Term THEN Rule_Hearing
  IF concept [Ec07] or word [馨|蘭|麝|蕙|檀] in Term THEN Rule_Smell
  IF concept [Br12|Bh05|Bh07] or word [杯|醉] in Term THEN Rule_Taste
  IF concept [Eb26|Eb27|Eb10|Eb13|Ea06] in Term THEN Rule_Touch
// Layer 2 Rule Set
Rule_Sight:
  IF Sense_Word == “金” THEN Rule_Golden
  IF Sense_Word == “紅” THEN Rule_Red
  IF...THEN...
Rule_Hearing:
  IF Sense_Concept == “Bi” THEN Rule_Animal
```

```

IF...THEN...
Rule_Smell:
.....
Rule_Taste:
.....
Rule_Touch:
.....

// Layer3 Rule Set
Rule_Golden:
  IF Object_Concept == "Bq(衣物)" THEN add 0.30 into Scenario_A, Add 0.24 into
    Scenario_B, Add 0.30 into Scenario_C, Add 0.16 into Scenario_D
  IF...THEN...
Rule_Red:
.....
Rule_White:
.....

// Diagnose/Output
IF MAX{Scenario_A,..., Scenario_H} == "Scenario_A" THEN output "浪漫的情愛"
IF MAX{Scenario_A,..., Scenario_H} == "Scenario_B" THEN output "女性的姿態"
IF MAX{Scenario_A,..., Scenario_H} == "Scenario_C" THEN output "怨曠與傷逝"
.....

```

第四章 系統實作與實驗

本章節藉由一些實驗設計與驗證來評估此專家系統對於判斷詞風與情境的成果，以及實驗成果的分析。

4.1 評估指標

對於系統的評估常使用召回率與精確度這兩個指標，其定義如下：

$$\text{召回率: (Recall)} = N3/N1 \quad (1)$$

$$\text{精確度: (Precision)} = N3/N2 \quad (2)$$

N1: 正確的詞風數目

N2: 專家系統辨認出的詞風與情境

N3: 專家系統正確辨認出的詞風與情境

召回率是指宋詞詞風與情境判斷專家系統正確辨認出的宋詞數目占正確的數目之比率，精確度是指在所有的宋詞數目當中，宋詞判斷專家系統正確辨認出的宋詞數目所占的比率。當召回率與精確度的值愈高時，表示宋詞專家系統的品質愈好。通常召回率和精確度不能兩全其美，若召回率上升，則精確度就會下降。這是因為在每次的判斷程序中，N1 的數目是固定不變的，為了要使召回率增高，代表判斷的宋詞數N2必需要增多。一旦N2 變多，就(2)而言，代表分母變大了，因此反而會使精確度下降。

此外，我們利用召回率與精確度的調和平均數 (Harmonic mean) 來判斷資訊的有效性 (Effectiveness) 定義如下：

$$\text{效度 (Effectiveness)} = 2PR / (P+R) \quad (3)$$

R：召回率

P：精確度

效度的數值愈高，代表取得的資訊愈有意義。我們以宋朝的婉約集與豪放集中的

一百闕詞為測試資料，以召回率、精確度和效度三個指標來評估專家系統的運作。

4.2 系統建置與實驗設計

宋詞的分析過程如圖 11 所示，首先將詞作經由宋詞斷詞器斷詞，再將斷詞結果詞彙經由詞風推論引擎作推論，判斷正確的詞風後再依所屬詞分進行情境推論。為了進行宋詞判斷實驗，我們實作了詞風與情境判斷專家系統。圖 12 為系統首頁，使用使可點選欲分析的選項；我們以分析詞風與情境選項為例，點選後如圖 13 所示，在畫面上方為輸入斷詞結果以及所屬詞牌，按下確定鍵後即開始分析詞風與情境，參考圖 14 及 15 所示。

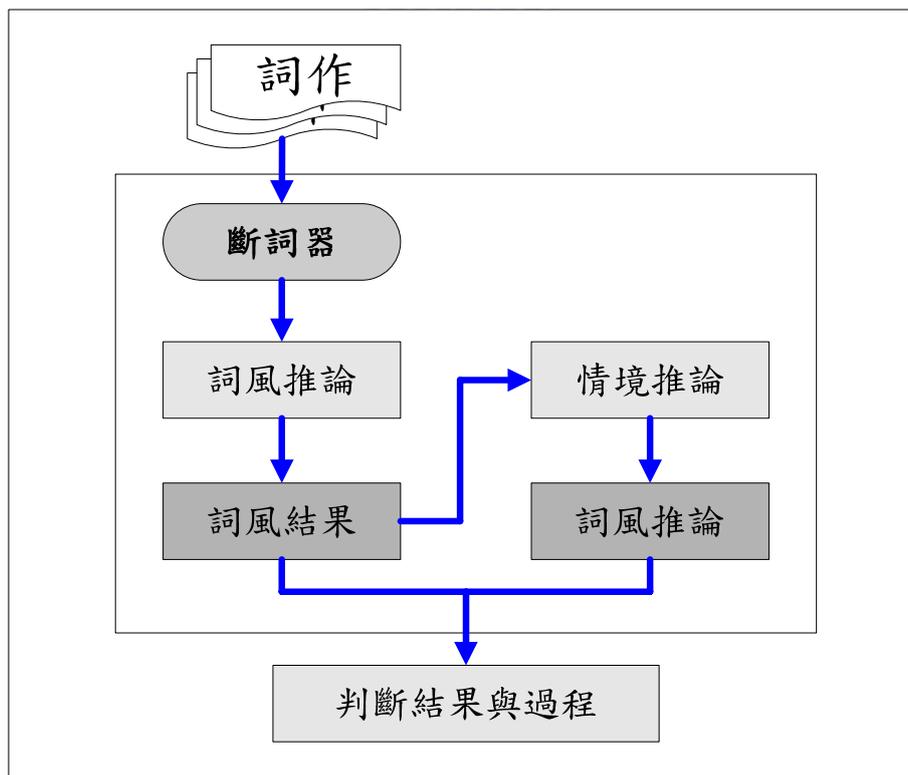


圖 11 詞風與情境判斷專家系統架構圖



圖 12 系統首頁



圖 13 輸入斷詞結果



圖 14 詞風判斷



圖 15 情境判斷

限於人力關係本實驗擷取【婉約集】與【豪放集】裡詞作中共 100 闕，婉約詞與豪放詞各佔一半，取其中的 50 闕當作建置「宋詞概念階層」的建置資料 (construction data)，其餘的則當作測試資料(testing data)。

4.3 詞風判斷實驗

在詞風判斷實驗部份，可區分為單以詞彙語意角度判斷以及詞彙語意加上節奏的角度判斷。

4.3.1 詞彙語意判斷

此實驗設計只針對由詞彙所判斷的語意作為詞風判斷結果的觀察，也就是說當透過形式數量規則集、內容特徵規則集、隱含特徵規則集所判斷的詞風為中性詞時，不作進一步根據詞調節奏的判斷。而實驗所得的判斷結果如表格 16 所示。

表 16 詞風判斷結果

	百分比
召回率 (Recall)	76 %
精確度 (Precision)	88 %
效度 (effectiveness)	81.6%

4.3.2 判斷語意判斷與節奏判斷

此實驗設計針對當如果詞風透過語意角度判斷為中性詞時，會進一步根據詞調節奏的特徵去作詞風的判斷。而實驗所得的結果判斷如表格 17 所示。由實驗觀察得知，加入詞調作為判斷規則時，反而使得準確度略為下降，這是因為有些作者在作詞時有時並沒有考慮節奏特性，例如，王惠青《滿江紅·太液芙蓉》是一闕婉約詞，但也使用【滿江紅】這種屬於豪放的詞牌。所以由實驗也讓我們得知分析詞作中的語意，才是判斷詞風的較佳方式。

表 17 詞風判斷結果

	百分比
召回率 (Recall)	78.8 %
精確度 (Precision)	83.4 %
效度 (effectiveness)	81.0%

由於詞彙的判斷結果不一定百分之百的正確，所以未來希望在系統的設計上保留專家校正的彈性，專家可重新設定詞彙的詞風。另外，如果詞彙不存在於《宋詞概念階層》也可手動加入，以擴增概念階層的完整性。

4.4 情境判斷實驗

完成詞風判斷之後，根據詞風結果判斷所屬的情境，婉約詞風下的情境可分為：浪漫的情愛、女性的姿態、怨曠與傷逝、離情與別恨、遊仙、詠物；而豪放詞風下的情境則分為：入世的政治態度、出世的理想抱負、邊塞與地方風物、詠物。我們根據文學專家知識建置判斷情境的規則，又利用花間集裡所統計感官對於每各情境使用的次數當作規則的 weight 值，所以實驗可區分為單就規則判斷以及加入 weight 判斷兩種情況。

4.4.1 情境判斷：單就規則，單一最佳解實驗

首先，我們以單就規則的方式，發現所得到的最佳解往往不只一個，所以我們初步定義當系統算出唯一的答案是為正確答案時，方為一個成功的案例，若算出兩個以上的答案則是為無法判斷。其所得出的結果相當令人失望，表 18 是為其召回率、精確度及效度：

表 18 情境判斷結果(單就規則-唯一最佳解)

	百分比
召回率 (Recall)	24.2%
精確度 (Precision)	8.4%
效度 (effectiveness)	12.4%

4.4.2 情境判斷：單就規則，多選解實驗

我們試著以不同的角度來分析系統給我們的資訊，既然系統給我們多個答

案，我們就把這些答案當作候選解答，也就是說，只要正確答案在候選解答裡就算是一個成功的案例，其所代表的意義是系統能夠提供使用者幾個可能答案的建議，讓使用者能夠從較少數的候選解答中挑出正確解。表 19 是為其召回率、精確度及效度：

表 19 情境判斷結果(單就規則-多數候選解答)

	百分比
召回率 (Recall)	80.5%
精確度 (Precision)	43.5%
效度 (effectiveness)	56.5%

4.4.3 情境判斷:加權值(Weight)實驗

接下來，我們把之前所提到的 Weight 導入規則集當中，當每一條規則在成立時，對於各種不同的情境會給予不同的加權，而在最後去計算一關詞符合多少規則並且總共給予每個情境多少的分數，取其最大者就是我們的最佳解。表 20 是為其召回率、精確度及效度：

表 20 情境判斷結果(加入 Weight)

	百分比
召回率 (Recall)	59.1%
精確度 (Precision)	67.9%
效度 (effectiveness)	63.2%

4.4.4 分析比較

綜合分析比較上述的實驗結果我們可以發現幾個現象：首先，使用單就規則的方式嘗試去找出一關詞作的情境基本上效果非常有限，無論是召回率、精確度或是效度都低得難以接受，其主要原因在於用這種方法所找出的答案往往不是單一解，使得算出的結果往往是“無法判斷”。但是這並不代表單就規則的方式無法

提供使用者任何有用的資訊，進一步分析我們的實驗結果可以發現，那些正確的情境往往包含在那些有一個以上答案的結果裡面，換句話說，如果系統是以提供使用者可能的情境的角度來看，其召回率大大提升至 80.5% (表格 19)，但是由於其提供的解答往往不只一個，所以其精確度依然低於 50%，相對的，其效度也無法超過 60%。

當我們運用加入 Weight 的方式去判斷時，所得到的結果就比較能夠讓人接受，用這種方式所算出的解答往往是單一的最佳解，根據表格 20 的數據我們可以發現，雖然其召回率下降至 60% 左右，但是其精確度及效度都大大提升至 60% 以上，對於一個從八種情境中判斷詞作屬於哪一個情境的問題而言，這樣的方式提供了相當大的協助。接下來我們的實驗將結合本系統中詞風判斷的機制，來探討詞風的預先分類能夠提升多少情境判斷的召回率、精確度及效度。

4.5 詞風與情境判斷實驗

上述兩個實驗室分別對詞分與情境個別作判斷的結果，接下來實驗則是結合上述兩種系統，做完整的判斷，亦即先將詞作作風格分類，然後再依據所屬風格判斷可能的情境。實驗結果如表 21、22 所示。

表 21 詞風與情境判斷結果(單就規則)

	百分比
召回率 (Recall)	86.4%
精確度 (Precision)	50.1%
效度 (Effectiveness)	63.4%

表 22 詞風與情境判斷結果(加入 weight)

	百分比
召回率 (Recall)	72.2%

精確度 (Precision)	76.8%
效度 (Effectiveness)	74.4%

比較表格 21 和之前的表格 19 我們可以發現，如果我們經由系統中的風格判斷機制事先得知詞作的風格，無論其召回率、精確度及效度都有所提升。再比較表格 22 和上一節的表格 20，我們可以看到無論在召回率、精確度或效度各方面都有十個百分點左右的提升而達到 70% 以上。證明我們在本論文當中所提出使用人類五感來判斷詞作所屬的情境，並且使用風格分類來提高準確度的方法，在實作上是可行的方案。



第五章 結論與展望

借助科技的進步，我們可將資訊融合中國古典文學，進而賦予古典文學新的生命，也藉由網路的普及讓更多人可以欣賞到詩詞的優美。詩詞對於情感的描述，不是直接的而是透過曲折的方式來表達，但也因為如此，想利用電腦來表達詩詞中的情感就更為困難。因此，我們透過擷取專家知識的方法，建構能判斷詞風與情境的專家系統，企圖讓電腦能夠像人類一樣，能夠去分辨一首詞是慷慨激昂或是委婉纏綿，以符合宋詞分類當中所普遍認同的豪放婉約詞風分類。另外，我們利用詞作裡所使用相對於人類五感的感官詞，來判斷此詞作所屬的情境。

在我們所架構的系統當中，針對構成詩詞的最小語意單位——詞彙作分析，用各種方式去區分其涵義為帶有豪放或婉約意味，進而用以判斷詞彙所構成的宋詞是屬於豪放或婉約。我們根據事物的形式數量、所描寫的情景以及專家無法分類但是在詞作運用上有特殊風格習慣的詞彙，建立了詞風概念階層並且根據它分別設計形式數量規則集、內容特徵規則集和隱含特徵規則集。此外，亦根據音調句法的節奏來設計詞調特徵規則集。最後，我們再統合以上各種種所得的資訊，建立判定規則，用以決定一闕宋詞作品的詞風屬於豪放或是婉約。在情境方面，我們根據洪華穗所著之“花間集的主題與感覺”一書所整理出的專家知識，利用詞作當中所使用的感官詞分佈，架構出情境概念階層，並且依此建立情境判斷規則集，來判斷詞作所屬的情境。

在系統實作及實驗方面，我們架構出一個詞風與情境判斷專家系統，利用部分的花間集以及宋詞集詞作來訓練我們所架構的系統，然後用另一部份的詞作做為我們的測試組，在風格判斷方面，我們得到 80% 左右的正確率，在情境判斷方面我們的正確率也都在 70% 以上，相較於即使以人工判斷亦有所出入的情況下，我們利用電腦計算出這樣的成果，無論在可行性或準確率上，皆可作為重要參考依據。

在現階段由於詞風分類上比較公認的仍然以豪放婉約為主，一種近似於二分法的方式，在本論文當中，我們更進一步地將它分細成八種情境，然而無論在各方面亦存在著相當的進步空間，在未來的工作裡，我們將著手於以下的方向：

1. 持續擴增並修訂所建立的各種詞風判斷資料庫，讓現有的詞風判定有更豐

富的參考依據，提高正確率。

2. 目前每一各概念對於判斷詞風的權重皆為相同，未來可透過 ML 方式調整概念的不同權重。並且設計依情況調整兩種風格的臨界值(並非永遠為 0)
3. 改進系統使其具有機器學習能力，使其能夠自動新增規則並且自我調整每條情境規則的情境加權值(Weight)。
4. 導入 Fuzzy 的觀念到我們在本篇論文所架構的各種規則集當中，使得系統在資訊不足的情況下依然能做有限度的推論

長遠目光來看，我們希望能夠讓電腦也能夠像人一樣去感知情緒，由於詩詞等文學作品，主要反映作者的情緒反應，所以我們試著以預期人類心理為出發點，並且試著用更接近於人類情緒的五官感知為出發點，讓電腦能夠更接近人腦思考模式。



參考文獻

- [1] B. Chandrasekaran, J. R. Josephson and V. R. Benjamins, “What are ontologies, and why do we need them?,” *IEEE Intelligent Systems*, 14(1): 20-26,1999.
- [2] M. Fernandez, A. Gomez-Perez and N. Juristo, “Methontology: From Ontological art towards ontological engineering,” *In Proceedings of Workshop on Ontological Engineering: AAAI-97 Spring Symposium Series*, Stanford, CA, 1997.
- [3] T. Gruber, Ontology definition,
<http://www-ksl.stanford.edu/kst/what-is-an-ontology.html>
- [4] M. Gruninger and Mark S Fox, “TOVE methodology for the design and evaluation of ontologies,” *Department of Industrial Engineering University of Toronto*, Toronto, Canada, M S A, 1995.
- [5] I. Niles and A. Pease, “Linking lexicons and ontologies: mapping WordNet to the suggested upper merged ontology,” *Proceedings of the IEEE International Conference on Information and Knowledge Engineering (IKE 2003)*, Las Vegas, Nevada, June 23-26, pp.412-416, 2003.
- [6] C.J. Tsai and S.S. Tseng, “Building a CAL expert system based upon two-phase knowledge acquisition,” *Expert Systems with Applications*, 22, 235-248, 2002
- [7] C.J. Tsai, S.S. Tseng and Y.C. Wu, “A new architecture of Object-Oriented rule base management system,” *Proc. of Tools Asia’99*, Nanjing, China, September 1999,.
- [8] M. Uschold and Michael Gruninger. *The knowledge engineering review*, 1996.
- [9] Y. Yi, Z.S. He, L.Y. Li, T. Yu and E. Yi, “Advanced studies on traditional Chinese Poetry style identification,” *Proceedings of the Fourth International Conference on Machine Learning and Cybernetics*, Guangzhou, China, 18-21, August 2005
- [10] 王迺仁、曾憲雄、楊哲青、蘇俊銘、羅鳳珠，“詩風規則之研究-以唐朝近體詩為例”，第二屆文學與信息科技國際研討會，北京大學，2005，12月8-10日。
- [11] 王迺仁，“唐詩之詩風探勘”，國立交通大學，碩士論文，2006年6月。

- [12] 谷聞，豪放詞。西北大學出版，1994。
- [13] 谷聞，婉約詞。西北大學出版，1994。
- [14] 余曉峰、劉鵬遠、趙鐵軍，“一種基於《知網》的漢語詞語詞義消歧方法”，第二屆全國學生計算語言學研討會，北京語言大學，2004年8月26日-27日。
- [15] 克萊夫·貝爾(英國著名視覺藝術評論家)，“藝術”，北京:中國文藝聯合出版社，1984。
- [16] 李良炎，“基於詞連接的詩詞風格評價技術”，重慶大學計算機學院，博士論文，2004年8月。
- [17] 洪華穗，花間集的主題與感覺，文津出版社，1999。
- [18] 梁榮基，詞學理論綜考，北京大學出版社，1991。
- [19] 許薰尹，“宋詞斷詞器與本體論之建置”，國立交通大學，碩士論文，2006年6月。
- [20] 梅家駒等編著，同義詞詞林，台灣東華書局股份有限公司，1997，12月。
- [21] 曾憲雄、黃國禎，人工智慧與專家系統理論/實務/應用，臺北市旗標，2005。
- [22] 羅鳳珠，“【倚聲填詞】格律自動檢測索引教學系統”，第六屆國際電腦輔助教學研討會論文集:頁432-438，1997，3月13-15日。
- [23] 羅鳳珠，“【倚聲填詞】格律自動檢測索引教學系統”。
- 取自<http://cls.admin.yzu.edu.tw/FillODE/default.htm>，1997，6月9日。
- [23] 蘇豐文、傅怡婷、陳書磊、楊世堯、羅鳳珠，“漢語詩的本體知識與語意檢索”，第一屆文學與資訊科技國際會議，清華大學，2003，12月9-11日。