

## 部落客憂鬱傾向分析與預測

### Analysis and Prediction of Blogger's Depression Tendency

董家銘 Chia-Ming Tung

國立成功大學資訊工程學系

Department of Computer Science and Information Engineering

National Cheng Kung University

[p7897127@mail.ncku.edu.tw](mailto:p7897127@mail.ncku.edu.tw)

盧文祥 Wen-Hsiang Lu

國立成功大學資訊工程學系

Department of Computer Science and Information Engineering

National Cheng Kung University

[whlu@mail.ncku.edu.tw](mailto:whlu@mail.ncku.edu.tw)

#### 摘要

憂鬱症已列為聯合國世界衛生組織視為新世紀三大疾病，與癌症、愛滋病一起蠶食鯨吞著人民的身心健康。根據聯合國世界衛生組織估計，全球目前有二到四億人口正為憂鬱症所苦，估計在亞洲至少約有五千萬的憂鬱症患者，且人數不斷上升。2020年，憂鬱症將與心臟病，成為影響人類生活甚巨的前二大疾病。根據統計，台灣地區2007年統計結果，憂鬱症盛行率約8.9%，換言之，超過兩百萬人罹患憂鬱症。40%的憂鬱症患者會有輕生或自殺的念頭，10~15%的患者因自殺而死亡。所以有效的找出有憂鬱傾向的民眾已經是一項不容忽視的醫療衛生議題。因此本研究提出一項創新的憂鬱傾向預測技術，利用部落格網誌文章自動判別部落客作者的憂鬱傾向。

隨著 Web 2.0 社群網路(Social Network)快速興起，使用者每天在部落格寫下工作和生活的諸多苦惱與需求，雖然已有許多部落格作者的情緒分析研究，但是目前並無相關研究開始探究部落格作者的憂鬱傾向預測，本研究參考美國精神科醫學會所發表之精神疾病診斷與統計手冊第四版修訂版(DSM-IV-TR) [1]中對重度憂鬱症的定義及從部落格作者的網誌文章分析，定義出四個憂鬱傾向的因素，其中包含事件、負面情緒、症狀和負面想法。然後利用這四項因素協助判斷部落格作者的憂鬱程度，因此本研究嘗試探究下列兩項重要研究議題，並開發相關處理技術：(1) 憂鬱症患者網誌文章的憂鬱傾向與相關因素分析，(2) 部落格作者的憂鬱傾向預測模型。

#### Abstract

According to the investigation report of the Department of Health, Executive Yuan, R.O.C. in 2007, it is estimated that about 7.3% of Taiwan's population suffer from the major depressive disorder. How to identify patients with depression tendency is one of important health issues. Thus, this project tries to develop a novel technique to automatically identify the depression tendency of bloggers using their blog posts.

With the fast growth of social networks, bloggers usually write daily posts with their emotion and events happened in work, home, or life. Although there are lots of research works about emotion analysis and classification, to our knowledge, there is no work focusing on prediction of blogger's depression tendency based on emotion analysis. In this project, we try to analyze key factors affecting major depressive disorder, such as negative event, negative emotion, symptom and negative thought, and then use these four factors to assist bloggers to predict depression tendency. Therefore, we focus on the investigation of the following two research issues (1) analysis of relevant factors of depression on blog posts written by patients with the major depressive disorder, (2) development of event-emotion-driven depression tendency prediction model.

關鍵詞：憂鬱傾向，事件，負面情緒，症狀，負面想法

Keywords: Depression Tendency, Event, Negative Emotion, Symptom, Negative Thought.

## 一、緒論

雖然台灣近年經濟起飛，生活富裕，但民眾似乎愈活愈不快樂，台灣民眾的生活快樂指數在東亞七個國家敬陪末座 [24]。近年來自殺的報導逐年增加，甚至許多年輕學子的自殺事件也時有所聞，這些自殺事件大部分是因為生活中遭遇巨大壓力，或是已經患有憂鬱病症一段時間，最終因強烈的負面情緒引發自殺負面想法而造成悲劇。憂鬱症已列為聯合國世界衛生組織視為新世紀三大疾病，與癌症、愛滋病一起蠶食鯨吞著人民的身心健康。根據聯合國世界衛生組織估計，全球目前有二到四億人口正為憂鬱症所苦，估計在亞洲至少約有五千萬的憂鬱症患者，且人數不斷上升。2020年，憂鬱症將與心臟病，成為影響人類生活甚巨的前二大疾病。根據統計，台灣地區2007年統計結果，憂鬱症盛行率約8.9%，換言之，超過兩百萬人罹患憂鬱症。40%的憂鬱症患者會有輕生或自殺的念頭，10~15%的患者因自殺而死亡。因此，憂鬱症的防治變得日益重要，所以有效的找出有憂鬱傾向的民眾已經是一項不容忽視的醫療衛生議題。

網際網路的快速發展造就社群網路的興起，許多社群網路服務網站提供人們彼此更快速方便的聯絡溝通模式，像是部落格(Blog)和近年火紅的微網誌(Micro Blog)。痞客邦(PIXNET)、隨意窩(Xuite)、yam天空部落等都是國內知名的部落格服務網站。另外臉書(Facebook)、推特(Twitter)、噗浪(Plurk)等都是非常熱門的國際微網誌服務網站。部落格提供網路使用者隨意撰寫文章紀錄生活中遭遇的點點滴滴，並抒發心情感受，有驚、有怒、有喜樂也有悲傷。根據我們對大量的部落格和論壇文章觀察，發現許多文章內容出現事件、負面情緒、症狀及負面想法等相關詞彙，如圖一所示，作者面對”重考”事件，心理產生多種負面情緒如”大哭”、”恐慌”、”焦慮”、”沮喪”等，結果也出現一些生理症狀”睡不好”，甚至負面想法”撐不下去了”。經由長期追蹤文章負面情緒詞的類型與出現頻率，可以了解部落格作者所發生的心理問題。所以本研究首先嘗試利用部落格文章內容來分析作者的負面情緒，然後偵測部落格作者是否有強烈情緒不穩或憂鬱傾向，或甚至產生自殺企圖。我們期待開發有效的創新技術，從部落格文章判斷有憂鬱傾向的作者，進而協助這些作者預防或治療憂鬱症。本研究特別關注兩項重要研究議題，並開發相關處理技術：(1)憂鬱症患者網誌文章的憂鬱傾向與相關因素分析，(2)發展部

落格作者的憂鬱傾向預測模型。因此我們提出事件情緒驅動的憂鬱傾向預測模型(Event-Emotion-driven Depression Tendency Prediction Model)，藉由負面情緒、事件、症狀和負面想法特徵的分析，然後判別出部落格作者的憂鬱傾向。

早上才在網誌上說要努力振作，  
結果現在還是失敗了，  
而且做了很不好的事。  
媽媽打來，又是要說課業的事情，  
一定是因為英文被扣考了，  
完全沒勇氣接電話，更何況電話響的時候我正在旁邊大哭，為了一些也許根本沒  
什麼的事。想強迫自己痛完就不准去在乎那些小事情，也不准再讓男友負擔我的  
情緒，哭著哭著就停了，像下過雨的冬季，雨停了還是一樣溫冷，雨的味道有雨  
的空氣好像總是無法驅離。  
想好好準備重考，  
可是每天都沒心思讀書。  
昨天到學校去還書，  
覺得自己真的越來越嚴重了，一走出捷運站就開始感到恐慌，晚上的夜市好熱鬧，  
路旁擦身而過的人都讓我很緊張，一路走到學校真的是一大折磨，進了學校就更  
可怕了，感覺隨時會遇見同班同學，雖然遇見了也不怎樣，反正又不熟，  
但是還是覺得很焦慮，還完書又要再重複一遍這些複雜的心理煎熬，回到家覺得  
好沮喪。  
突然覺得我的未來要怎麼辦？書讀不好，工作也沒了。  
想到這裡就害怕繼續想下去，結論永遠都指著同一個方向，可是我不想這樣，  
但同時卻又無路可走。  
越慌張就越沒辦法讀書，想到曠課紀錄破表的學校也很痛苦，唯一能離開的方法  
就是認真讀書可是卻又沒心思。  
很多事不敢告訴男友。今天好不容易說了我想去看醫生，雖然沒有說實話，  
只說了我睡不好而已，我不敢一個人去醫院，就要等到下個月才能讓他陪我去，  
可是我覺得我快撐不下去了。

圖一、憂鬱症患者發表網誌文章（藍色矩形實線代表事件；紅色矩形虛線代表負面情緒；紫色橢圓實線代表症狀；綠色橢圓虛線代表負面想法）

## 二、 文獻探討

### （一） 憂鬱症簡介與臨床診斷技術

憂鬱症通常是指重性憂鬱障礙(major depressive disorder)，在醫學上被視作一種精神疾病。憂鬱症患者的典型症狀是心情低落，通常會沉浸在憂鬱的情緒狀態中，對一些有興趣的事物皆感無趣，絕望或是認為人生沒有價值，更甚者會有自殺念頭。在生理上也會出現症狀，例如失眠、沒有食慾造成體重下降、疲勞、無精打采沒有活力或是出現腸胃問題等。

憂鬱症的病因在醫學上目前尚未確認，從心理學的研究認為一個人的人格發展在許多方面促使了憂鬱症的發作，很多學派支持這個論點，例如精神分析學、存在主義心理學等[10, 12]。而社會學的研究觀點則前瞻地指出身處的環境、人際關係與遭遇的生活

事件是罹患憂鬱症的重要原因[26]，例如家庭功能受損或是處於惡劣工作環境中。本研究認為生物、心理、社會這三個因素都是重要影響因素，另外對於影響部落格作者負面情緒的負面生活事件將深入探究。

現今醫學的憂鬱症診斷標準主要是根據美國精神疾病協會(American Psychiatric Association, APA) 的精神疾病診斷與統計手冊第四版修訂版(DSM-IV-TR) (2000)[1]和世界衛生組織(World Health Organization, WHO) 的國際疾病與相關健康問題統計分類(ICD-10) (2007) [17]，另外董氏基金會[25]也有提供憂鬱症自我診斷評量表讓一般民眾自我篩檢。下面簡單說明 DSM-IV-TR 有關憂鬱症的九項診斷標準：

1. 憂鬱情緒：快樂不起來、煩躁和鬱悶。
2. 興趣與喜樂減少：提不起興趣。
3. 無法專注：無法決斷、矛盾猶豫、無法專心。
4. 體重和食慾失常：體重下降或增加、食慾下降或增加
5. 失眠(或嗜睡)：難入睡或整天想睡。
6. 精神運動性遲滯(或激動)：思考動作變緩慢、腦筋變鈍。
7. 疲累失去活力：整天想躺床、體力變差。
8. 無價值感或罪惡感：覺得活著沒意思、自責難過，都是負面想法。
9. 自殺意圖：反覆想到死亡，甚至有自殺意念、企圖或計畫。

## (二) 部落格情緒分析

近幾年來部落格的服務快速崛起，許多學者發現情緒是部落格網誌文章中一個重要的成分，而國內台大資工系陳信希教授最近幾年也有不少的研究關注在部落格的情緒分析上[11, 22, 23]。此外，近幾年以情緒偵測和分類為主的相關研究[11, 20]快速增加，而部落格網誌內容主題分析也逐漸地變成熱門的研究對象[6]。許多研究[3, 11, 19, 20, 21]皆針對部落格網誌文章的內容主題與情緒，提出各式各樣的方法進行情緒辨識與分類。Leshed and Kaye[9]針對LiveJournal8.com 網站部落格使用者如何詮釋他們的心情做了一個全面性的調查，該網站提供了 132 種心情選項當作部落格使用者情緒標記。Hsu 和 Lin[3]也提供了一個 SVM 分類器對大量的文字列表做心情的關聯分析。例如，”電腦”(Computer)這個詞彙相當有可能被歸類到”煩悶”的部落格條目。Yang 等人[19, 20, 21]利用 SVM 和 CRF 對部落格網誌文章作情緒的分類，他們以句為單位和以文章為單位進行大規模效能評估。這些情緒偵測和分類的相關研究觸發本研究對於情緒研究的延伸應用，我們嘗試利用負面情緒特徵提出創新的部落格作者的憂鬱傾向預測研究。

## (三) 部落格事件擷取

本研究是以紀錄部落格作者個人日常生活上引發負面情緒的負面事情作為事件的定義，大概以家庭、感情、學業、工作四種類型為主的生活事件。過去有關事件擷取的相關研究主要來自於兩個研究領域，一個是主題偵測與追蹤(Topic Detection and Tracking, TDT)，另一個是自然語言處理。

Chen 等人 (2008)提出的 TSCAN [23]和 Kumaran 等人[7]提出的 NED 皆是 TDT 相關的研究，在此他們對於 Topic 的定義是一個有重大影響和意義的事件或活動。大部份有關事件的相關研究主要都是以新聞熱門事件為主[8]。Teng and Chen [22]提出了利用 Temporal Collocation 的方法對部落格網誌文章做事件抽取，但是他們著重在熱門主題事件抽取，相對來說本研究則深入探究更瑣碎的生活事件抽取為主，技術上應該比較困難。

這幾年在自然語言處理領域中，有需多相關事件擷取的研究。2005 年 Pustejovsky

[14, 15]為了探究事件和時間的關係提出 TimeML (Time Markup Language)概念。接著 2007 年，Mani (2007)利用 TimeML 的時間標註結構，標註新聞文章的事件、時間和彼此關係。Pustejovsky 開始進行 Textual Inference Tasks [4, 5, 16] 的研究，首先他以 TimeML 為基礎建構 Event Structure Lexicon (ESL)，針對 Event Implicature 和 Entailed Subevent 進行推論。Palmer 等[13]作了不少由動詞出發的事件表達，分析，與預測研究。他們對於 Event Relation 的偵測，主要是運用學術上廣泛使用的詞典資源，如：Wordnet 和 Framenet，來訓練包含語法與語意資訊的特定領域動詞詞彙網 Verbnet。

### 三、研究方法

為了從部落格作者的網誌文章瞭解和預測部落客作者的憂鬱傾向，本研究搜集了大量憂鬱症患者的網誌文章或 BBS 論壇文章，藉由深入觀察事件、負面情緒、症狀、負面想法等四項重要特徵對於憂鬱症的複雜影響與關聯，我們首先嘗試提出創新的分析和預測技術：(一) 提供憂鬱症患者網誌文章的憂鬱傾向與相關因素分析，(二) 建構一個事件情緒驅動的憂鬱傾向預測模型 (Event-Emotion-driven Depression Tendency Prediction Model)。這些分析報告和創新技術應該能夠有效地提早判斷具有憂鬱傾向的部落客作者，建議他們盡速尋求專業醫療的協助。下面我們將詳細說明本研究研究及發展的主要分析方法與預測技術。

#### (一) 憂鬱症患者網誌文章的憂鬱傾向與相關因素分析

##### 1. 初步的觀察和分析

根據我們對憂鬱症患者網誌文章的觀察，當憂鬱症患者在撰寫文章時經常會出現各種負面情緒字眼例如“大哭”、“恐慌”、“焦慮”、“沮喪”(圖一)，而這些負面情緒大部分由一般生活上的事件所引起，如“重考”。比較嚴重的憂鬱症患者在撰寫文章時，除了會出現比較強烈的負面情緒字眼，甚至會出現身體症狀和引發心理嚴重的負面想法，例如“想死”和“自殺”。

##### 2. 主要的分析方法：

藉由大量憂鬱症患者網誌文章觀察，我們提出以事件、負面情緒、症狀、負面想法四項重要特徵為主的憂鬱傾向創新分析方法，然後我們根據兩個醫學分析模型來探討我們提出的分析方法的有效性和優點。

##### (1) 生物心理社會分析模型

1977 年學者 Engel[2]提出生物、心理、社會三合一分析模型(Biopsychosocial Model, BPS Model)的新醫學概念，對於病患面對疾病的分析因素包含了生物面(Biological)、心理面(Psychological)和社會面(Social)的三方因素。其中生物面即是身體症狀(Symptom)，心理面則涵蓋了情緒(Emotion)、想法(Thought)與行為(Behavior)，社會面即是病患面對的環境因素，換個角度講，也就是病患所發生的生活事件(Event)。透過進一步的分析比較後，我們可以將 BPS 模型的三個主要因素對應到我們提出的創新憂鬱症分析方法的四項因素，如表一所示，這樣的相似對應關係顯示我們的分析方法應該全面性地涵蓋憂鬱傾向的相關重要因素。

表一、BPS model 與本研究分析方法的對應關係

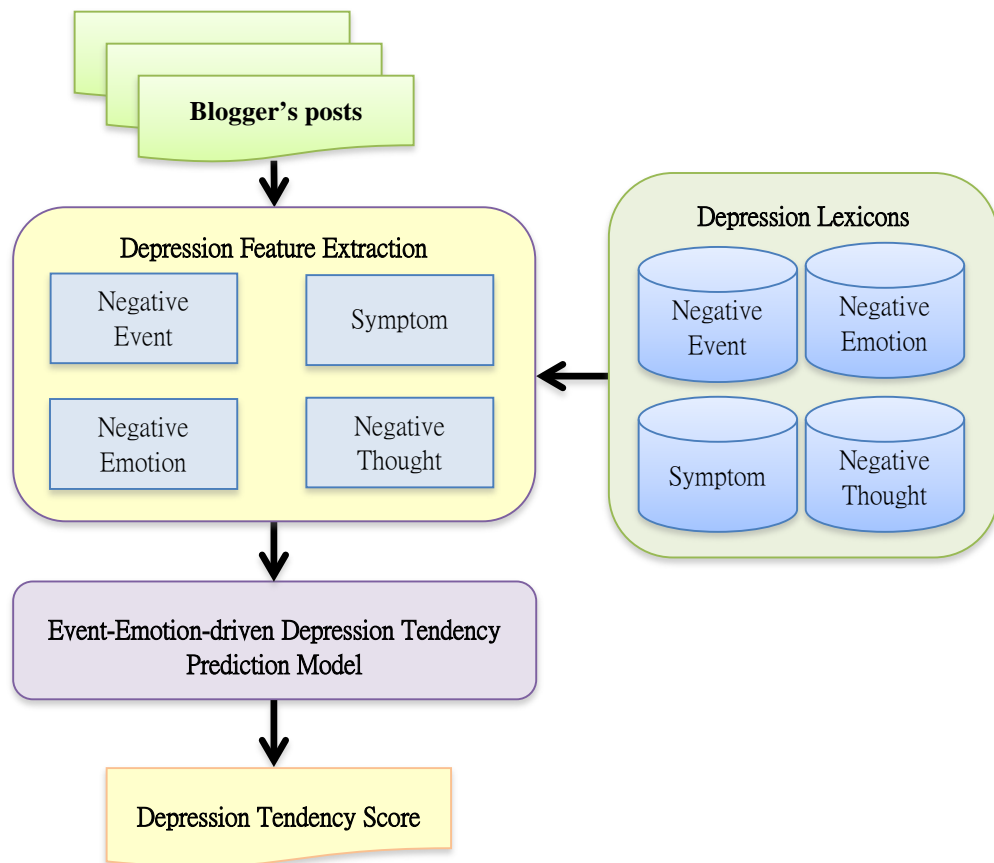
Analysis Model	Analysis Factors				
BPS model	Biological	Psychological			Social
Correspondence	Symptom	Emotion	Thought	Behavior	Event
Our analysis method	Symptom	Negative Emotion	Negative Thought		Negative Event

## (2) DSM-IV-TR 憂鬱症臨床判別標準

現今的醫學尚未提供生物檢測方法以直接確診憂鬱症患者。以重度憂鬱症而言，精神科醫師目前主要根據美國精神疾病協會(American Psychiatric Association, APA) 的精神疾病診斷與統計手冊第四版修訂版(DSM-IV-TR)的九項判別標準對病患做篩檢和診斷。如表二所示，我們提出的憂鬱傾向分析方法的負面情緒因素，可以相對應 DSM-IV-TR 的前三項判別標準(項次 1、2、3)。憂鬱症患者的網誌文章也常出現負面情緒伴隨著負面的生理症狀，譬如說哭泣、頭痛、失眠、食慾下降等等。這些症狀剛好符合表 1 中 DSM-IV-TR 的 4、5、6、7 項判別標準。另外憂鬱症患者的網誌文章也常出現負面想法，例如自殺、跳樓、自殘、燒炭等等。這些負面想法也符合表二中 DSM-IV-TR 的第 8、9 兩項判別標準。根據這三項因素的對應關係，可以清楚顯示我們的分析方法應該有效地涵蓋憂鬱症臨床診斷判別標準。值得注意的是我們提出的**事件因素**並未出現在 DSM-IV-TR 的九項判別標準，本研究提出這項創新因素，並深入探究是否可以有效地協助分析憂鬱傾向。

表二、DSM-IV-TR 重度憂鬱症九項判別標準與本研究提出的憂鬱因素對應關係

判別標準	說明	憂鬱傾向因素
1	憂鬱情緒：快樂不起來、煩躁、鬱悶	負面情緒
2	興趣與喜樂減少：提不起興趣	
3	無法專注：無法決斷、矛盾猶豫、無法專心	
4	體重和食慾失常：體重下降(或增加)、食慾下降(或增加)	症狀
5	失眠(或嗜睡)：難入睡或整天想睡	
6	精神運動性遲滯(或激動)：思考動作變緩慢、腦筋變鈍	
7	疲累失去活力：整天想躺床、體力變差	
8	無價值感或罪惡感：覺得活著沒意思、自責難過，都是負面想法	負面想法
9	自殺意圖：反覆想到死亡，甚至有自殺意念、企圖或計畫	



圖二、部落客作者的憂鬱傾向預測系統架構

## (二) 部落客作者的憂鬱傾向預測模型

### (1) 問題和構想

憂鬱原本是一個抽象概念，電腦系統無法從部落客作者文章中直接判別他的憂鬱傾向，然而根據我們對憂鬱症患者網誌文章的觀察，憂鬱症患者撰寫的文章經常會出現負面情緒、事件、症狀和負面想法（如圖一）。因此為了讓電腦系統有效地自動預測部落客作者的憂鬱傾向，我們構想提出**事件情緒驅動的憂鬱傾向預測模型 (Event-Emotion-driven Depression Tendency Prediction Model)**，藉由負面情緒、事件、症狀和負面想法特徵的分析，然後判別出部落客作者的憂鬱傾向。圖二展示我們提出的部落客作者的憂鬱傾向預測系統架構，首先輸入某位部落客作者單篇或數篇的部落格文章，利用四個憂鬱因素詞彙集，系統可以擷取出負面情緒、事件、症狀和負面想法四種類型的憂鬱特徵，然後使用事件情緒驅動的憂鬱傾向預測模型 (EEDTP) 來計算部落客作者的憂鬱傾向分數。

### (2) 事件情緒驅動的憂鬱傾向預測模型 (Event-Emotion-driven Depression Tendency Prediction Model)

給定一篇部落格文章  $b$ ，我們想要利用機率模型來估算這篇文章透露出的憂鬱傾向 (Depressive Tendency)  $D$  的強度：

$$P(D|b) \quad (1)$$

根據本研究觀察具有高度憂鬱傾向作者的部落格文章，發現作者寫下具有負面情緒 (Negative Emotion) 的部落格文章時，往往有很高的比例是因為某些事件 (Event) 造成作者心情不佳，或是伴隨著症狀 (Symptom)、更甚有不好的負面想法 (Negative Thought)。我們將負面情緒、事件、症狀及負面想法這四項稱為憂鬱因素，因此公式 (1) 的憂鬱傾向  $D$  轉化成憂鬱因素  $D_f$ ，變成公式 (2)：

$$P(D|b) = P(D_f|b) \quad (2)$$

消極因素  $P_f$  涵蓋了事件  $E$ 、負面情緒  $M$ 、症狀  $S$  及負面想法  $T$ ，因此改成公式 (3)：

$$P(D_f|b) = P(E, M, S, T|b) \quad (3)$$

在四個消極因素之中，本研究同時提出消極因素的根源來自於引發憂鬱情緒的事件，稱為事件 (Negative Event)，接著事件可能引發負面情緒、症狀及負面想法。將公式 (3) 針對消極因素中的事件與其他三個消極因素展開，即為下列公式 (4)：

$$P(D|b) = P(E|b)P(M|E)P(S|E, M)P(T|E, M) \quad (4)$$

公式 (4) 即是本研究的創新模型，稱做事件情緒驅動的憂鬱傾向預測模型 (Event-Emotion-driven Depression Tendency Prediction Model)，而每篇部落格文章透過此模型得到的分數稱之為憂鬱傾向分數 (Depression Tendency Score)。

### (3) 憂鬱傾向分數估算

為了計算彈性方便，本研究利用 Log Linear Model 找出  $P(D|b)$  的最大機率值，公式如下：

$$P(D|b) = \frac{1}{Z} * \exp \sum_{f_i \in F} \lambda_i f_i(E, M, S, T) \quad (5)$$

公式 (5) 中的  $Z$  是數值轉機率的正規因素 (normalization factor)， $\lambda_i$  為權重係數， $f_i$  為特徵函數， $F$  為特徵函數集合， $F = \{f_{Ev}, f_{Ev-Em}, f_{sym}, f_{NT}\}$ 。我們將  $f_{Ev}(E, b)$  稱做事件特徵函數； $f_{Ev-Em}(E, M, b)$  稱做事件和情緒配對特徵函數； $f_{sym}(S, E, M, b)$  稱做症狀特徵函數； $f_{NT}(T, E, M, b)$  稱做負面想法特徵函數。

在接下來的小節會詳細介紹這五個特徵函數的計算方法。

- **事件特徵函數：**

經由比對部落格文章內容的詞彙  $E$  與憂鬱事件詞典  $E_L = \{e_1, e_2, \dots, e_n\}$ ，若詞彙  $E$  出現在事件詞典  $E_L$  中則給予事件詞分數。我們由台大 BBS PTT 的 Prozac 板 (憂鬱板) 中 378 篇已被專家標記成有憂鬱傾向的文章裡找出每個引發憂鬱的事件詞  $e_i$  在這 378 篇文章的詞頻  $freq(e_i)$ ，以詞頻的比例來當成該事件詞的重要程度，因此事件詞彙  $E$  可由公式 (6) 來計算。

$$f_{Ev}(E, b) = \frac{freq(E)}{\sum_{E \in E_L} freq(E)} \quad (6)$$

- **事件-情緒特徵函數：**



從 378 篇被專家標記成有憂鬱傾向的文章中，找出及事件  $E$  跟負面情緒  $M$  的配對，每一個負面情緒詞  $M$  都有情緒強度  $Intensity(M)$ ，情緒強度值介於 0 到 1 之間。我們首先要找出每一個負面情緒詞彙和事件詞彙配對相隔的距離  $dis(E, M)$ ，然後以常態分配 (Normal distribution) 機率來代表事件和情緒配對關係值，接著乘上該負面情緒強度值，因為一個事件可能引發多種情緒，因此加總所有事件情緒配對分數如公式(7)：

$$f_{Ev-Em}(E, M, b) = \begin{cases} \sum_M Normal(dis(E, M)) * Intensity(M), & \text{if } M \text{ exists} \\ 0, & \text{otherwise} \end{cases} \quad (7)$$

- **症狀特徵函數：**

如果一篇含有事件詞彙的文章，出現一些症狀詞彙，表示作者可能有明顯的憂鬱傾向，為了避免誤判，我們仍想要找出事件、負面情緒和症狀三者有強烈關係的結構，但計算上為了避免 sparsity 問題，我們以事件症狀配對，以及情緒症狀配對同時計算如公式(8)：

$$f_{sym}(S, E, M, b) = \sum_S \frac{Intensity(M) * (dis(E, S) + dis(M, S))}{2 * dis(E, S) * dis(M, S)} \quad (8)$$

- **負面想法特徵函數：**

和上一小節的症狀特徵函數相同的計算方法，如果一篇有事件詞彙的文章，出現了負面想法詞彙，表示作者可能有強烈的憂鬱傾向，我們想要找出事件、負面情緒和負面想法三者有強烈關係的結構，但計算以事件負面想法配對，以及情緒負面想法配對同時計算如公式(9)：

$$f_{NT}(T, E, M, b) = \sum_T \frac{Intensity(M) * (dis(E, T) + dis(M, T))}{2 * dis(E, T) * dis(M, T)}$$

(9)

## 四、 實驗

### (一) 資料集(Data Set)

本研究為了分析高憂鬱傾向的網路使用者文章，蒐集 PTT Prozac 板(憂鬱板) 2004/3/22~2011/9/21 期間 3458 個作者，共 17,985 篇文章，挑選文章數量最多的前 30 作者文章，將這些文章以 2 個星期為一個週期去切割，把該期間的文章集合起來成為一個觀察資料集，稱為一個 Window Size，用表二的準則挑選符合憂鬱傾向的 windows(包含該 windows 的前後各兩個 windows)，總共 162 個 windows，724 篇文章。

### (二) 憂鬱文章標記

憂鬱傾向的判定是需要由專業人士來判定，而本研究實驗的文章標記是由 3 位心理學專家標記，主要分成單篇文章標記是否有無憂鬱傾向以及 2 週為一個單位使用 DSM-IV-TR 的準則來判定，每篇文章皆會被標記成有沒有憂鬱傾向以及符合哪些 DSM-IV-TR 準則，在 2 週內為單位的文章裡 DSM-IV-TR 的準則的標記超過 5 種準則時，則該期間(window)則判斷為有憂鬱傾向。最終文章憂鬱傾向判定結果採用多數決方式，

三位專家標記結果，以兩位以上認定的結果為該文章或區間的憂鬱傾向結果。

### (三) 憂鬱因素辭典

本研究利用上述收集標記憂鬱傾向的文章，人工建立了四種詞典，各辭典的詞彙數量如表三所示。

表三、各種憂鬱因素辭典的詞彙數量

	情緒	事件	症狀	負面想法
詞彙數量	1316	201	58	31

### (四) 效能評估

為了評估本研究所提出的 EEDTP 模型效能，我們採用 SVM 及 DSM-IV-TR 自動標記兩種實驗為比較基準。SVM 是以本研究所蒐集的憂鬱因素詞典中所有的詞彙為 SVM 中的特徵以取得 SVM 分類結果。DSM-IV-TR 自動標記則是採用表二所列的 9 項評估標準及重度憂鬱症判定準則來進行評估。EEDTP 則是利用 10 倍交叉驗證方法進行模型評估。

從表四的實驗結果來看，本研究所提的 EEDTP 模型 (事件情緒驅動的憂鬱傾向預測模型, Event-Emotion-driven Depression Tendency Prediction Model) 的精確率(Precision)、召回率(Recall)、F measure、正確率(Accuracy)都比 SVM 和 DSM-IV-TR 方法好。

表四、憂鬱傾向預測模型預測結果比較

	Precision	Recall	F -measure	Accuracy
EEDTP	0.593	0.668	0.624	0.585
SVM	0.565	0.541	0.552	0.572
DSM-IV-TR	0.463	0.555	0.504	0.451

### (五) 實驗分析

#### (1) 正面例子：

表五中，事件出現次數很少，但 EEDTP 模型找出的事件詞「閱讀障礙」分數很高，所以可以判斷正確。表六中，事件詞「傷害我」雖然與症狀詞、負面想法詞以及部分負面情緒詞的距離很遠，但因為事件詞分數很高，所以判斷正確。

表五、 EEDTP 模型判斷正確的正面例子(1)

我越來越擔心我是不是得了強迫症了。  
閱讀障礙、理解障礙、強迫記憶之後，接著來了強迫思考！  
 我會一直想著同一件事情！或想著同一個字，想著那個字該怎麼寫。  
 能不能有人告訴我這是怎麼回事！我要崩潰了！  
 我不要再讀文章了！但我就是犯賤！明明有強迫記憶，硬要讀人家的文章，偏偏我又有閱讀障礙！想問別人這是什麼意思，又卡在理解障礙(別人說的話我聽不懂)。  
 最後來個強迫思考，想著這句話這篇文章到底是什麼意思！  
 我真的只能卡在自己的世界裡了。(大哭)

事件詞	症狀詞	負面想法詞
閱讀障礙	無	無

表六、EEDTP 模型判斷正確的正面例子(2)

我也真蠢 困擾這麼久 太傻 太傻  
 今天 期待 你接我下班期待第一次和你的晚場電影哭泣...原來是美好的因為  
 哭一哭 讓我就會清醒太久不會哭了  
 所以終於 宣洩出內心的壓力 還有冷漠的聆聽聽到我內心真正要下定決心的  
 遠離那焦慮謝謝你  
 我也恨你 那些焦慮 帶給我的痛苦  
 你沒有資格再來我心裡折磨我  
 從今以後 我要趕離你  
 請你這焦慮 從此遠離我的生命 生活消失  
 真正需要平靜的我 要回到我正常的一切生活如果不是焦慮 我就不會胡言  
 亂語不會失眠的更嚴重  
 .....  
 但我好不捨 你這樣說 讓我愧疚萬分 對不起給你的這樣如此大的壓力  
 可是我才知道 你是那麼愛我  
 即便我不停的說 都是我害你的你仍不願再傷害我 直說不是我的錯  
 好傻 我就活在你幸福的包圍卻要莫名的害怕焦慮恐慌

事件詞	症狀詞	負面想法詞
傷害我	失眠	消失

(4) 錯誤例子：

在表七的錯誤例子中，專家標記本篇文章無憂鬱傾向，但 EEDTP 模型找出的事件詞為「病情」，這個詞在憂鬱詞典訓練的憂鬱分數很高，因為該詞彙在訓練資料中出現的詞頻較高，而導致「病情」在本篇文章擷取事件時的分數變高，進而判斷成有憂鬱傾向。

表七、EEDTP 模型判斷錯誤例子

<p>現在吵得正夯的全面啟動，讓我越來越不想去看。不是因為被雷到。                  (本來沒人在說的時候，很想去看的。)                  今天看診的時候，突然有種很想逃離醫院的感覺。希望他永遠不要叫到我。                  (本來剛開始，我很期待進診間談<b>病情</b>的事。)                  我好想要被關心，但是被同一個人過度關心會覺得<b>反感</b>。真是太噁心了。                  (從<b>陌生</b>到熟識，這樣的感覺應該很好啊！不是嗎？)                  我不是特別、我不是刻意和別人不一樣。                  我是<b>有病</b>！</p>		
事件詞	症狀詞	負面想法詞
病情	無	無

## 五、結論

本論文提出事件情緒驅動的憂鬱傾向預測模型(Event-Emotion-driven Depression Tendency Prediction Model)，藉由負面情緒、事件、症狀和負面想法四項憂鬱特徵的分析，然後判別出部落格作者的憂鬱傾向。實驗結果顯示本方法可以有效地利用部落格網誌文章自動判別部落格作者的憂鬱傾向。本研究發展的部落格作者憂鬱傾向分析與預測創新技術，應該可以協助憂鬱症患者提早診斷，進一步尋求專業醫療，減輕痛苦。為了進一步提升本方法效能，未來我們將收集和標記更多訓練資料，以及擴充四個類型憂鬱因素辭典，尤其是症狀及負面想法詞典必須增加到一定的數量。

## 參考文獻

- [1] American Psychiatric Association, Diagnostic and Statistical Manual of Mental Disorders, Fourth Edition - Text Revision (DSMIV-TR), Amer Psychiatric Pub, 2000.
- [2] G.L. Engel, "The need for a new medical model: a challenge for biomedical medicine." Science, New Series, vol. 196, no. 4286., pp. 129-136, Apr. 8, 1977.
- [3] Hsu, C.W. and Lin, C. J., A comparison of methods for multi-class support vector machines. IEEE Transactions on Neural Networks. 2002, Vol. 13 (2002), p. 415-425.
- [4] Im, S. and Pustejovsky, J.. Annotating Lexically Entailed Subevents for Textual Inference Tasks, in Proceedings of FLAIRS Conference. 2010.
- [5] Im, S. and Pustejovsky, J.. Annotating event implicatures for textual inference tasks. Generative Approaches to the Lexicon (GL2009). 2009.

- [6] Judit, B. I, An outsider's view on "topic-oriented blogging", in Proceedings of the 13th international World Wide Web conference on Alternate track papers. 2004, ACM: New York, NY, USA. pp. 28-34.
- [7] Kumaran, G. and Allan, J., Text classification and named entities for new event detection, in Proceedings of the 27th annual international ACM SIGIR conference on Research and development in information retrieval. 2004, ACM: Sheffield, United Kingdom. p. 297-304.
- [8] Kuo, J.J. and Chen, H.H., Multidocument Summary Generation: Using Informative and Event Words. *ACM Transactions on Asian Language Information Processing (TALIP)*, 2008. 7(1): pp. 1-23.
- [9] Leshed, G. and Kaye, J.J., Understanding how bloggers feel: recognizing affect in blog posts, in CHI '06 extended abstracts on Human factors in computing systems. 2006, pp. 1019-1024.
- [10] Freud, S., Strachey, J., Richards, A., *On Metapsychology: The Theory of Psychoanalysis*. Penguin Books, 1984, pp. 251-268.
- [11] Lin, H.Y. and Chen, H. H., Ranking reader emotions using pairwise loss minimization and emotional distribution regression, in Proceedings of the Conference on Empirical Methods in Natural Language Processing. 2008, pp. 136-144.
- [12] A. Maslow, *The Farther Reaches of Human Nature*. New York, NY, USA: Viking Books, 1971, pp.318.
- [13] Palmer, M., Hwang, J. D., Brown, S. W., Schuler, K. K., Lanfranchi, A., Leveraging Lexical Resources for the Detection of Event Relations, in Proceedings of the AAAI 2009 Spring Symposium on Learning by Reading. 2009, pp. 81-87.
- [14] Pustejovsky, James, Robert Ingria, Roser Sauri, Jose Gastano, Jessica Littman, Rob Gaizauskas, Andrea Setzer, Graham Katz, and Christopher Habel. *The Specification Language TimeML*, Oxford University Press. 2005a.
- [15] Pustejovsky, James, Kiyong Lee, Harry Bunt, and Laurent Romary. "ISO-TimeML: An International Standard for Semantic Annotation." LREC 2010, Malta. May 18-21, 2010. Pieter M.A., Hekkert, P., Special Issue Editorial: Design & Emotion, *International Journal of Design*. 2009, Vol. 3, No. 2, p. 1–6.
- [16] Pustejovsky, James, Robert Knippen, Jessica Littman, and Roser Saurí”, “Temporal and Event Information in Natural Language Text”, *Language Resources and Evaluation*. 2005b, Vol 39, p.123-164.
- [17] World Health Organization, “The International Statistical Classification of Diseases and Related Health Problems 10th Revision (ICD-10),” <http://apps.who.int/classifications/apps/icd/icd10online>, 2007.

- [18] World Health Organization, “Depression,”  
[http://www.who.int/mental\\_health/management/depression/definition/en](http://www.who.int/mental_health/management/depression/definition/en).
- [19] Yang, C. H., Kuo, H. A. and Chen, H. H., Building emotion lexicon from weblog corpora, in Proceedings of the 45th Annual Meeting of the ACL on Interactive Poster and Demonstration Sessions. 2007, pp. 133-136.
- [20] Yang, C. H., Kuo, H. A. and Chen, H. H., Emotion Classification Using Web Blog Corpora, in Proceedings of the IEEE/WIC/ACM International Conference on Web Intelligence. 2007, IEEE Computer Society. p. 275-278.
- [21] Yang, C. H., Kuo, H. A. and Chen, H. H., Emotion Trend Analysis Using Blog Corpora, in Proceedings of the 19th Conference on Computational Linguistics and Speech Processing. 2007. pp. 205-218.
- [22] Teng, C. Y., Chen H. H., Event Detection and Summarization in Weblogs with Temporal Collocations, in International Conference on Language Resources and Evaluation. 2008. pp. 197-200.
- [23] Chen, M. C., Chen H. H., TSCAN: A Content Anatomy Approach to Temporal Topic Summarization, in IEEE Transactions on Knowledge & Data Engineering. 2008. pp. 170-183.
- [24] 林思恩，「東亞七國快樂指數 香港倒數第二 台灣人最不快樂」，  
[http://www.gospelherald.com.hk/news/soc\\_1311.htm](http://www.gospelherald.com.hk/news/soc_1311.htm)，2009。
- [25] 財團法人董氏基金會，「憂鬱情緒自我篩檢」，  
<http://www.jtf.org.tw/psyche/melancholia/overblue.asp>。
- [26] Paykel, E.S., Meyers, J.K., Dienes, M.N., Klerman, G.L., Lindenthal, J.J., Pepper, M.P., “Life events and depression : A controlled study,” Archives of General Psychiatry, vol. 21, no. 6, pp. 753-760, 1969.