

基於意見詞修飾關係之微網誌情感分析技術

Microblog Sentiment Analysis based on Opinion Target Modifying

Relations

王正豪 Jenq-Haur Wang

國立台北科技大學資訊工程學系

Department of Computer Science and Information Engineering

National Taipei University of Technology

jhwang@csie.ntut.edu.tw

葉庭瑋 Ting-Wei Ye

國立台北科技大學資訊工程學系

Department of Computer Science and Information Engineering

National Taipei University of Technology

bad00124@gmail.com

摘要

如何有效分析文件的意見傾向，一直是熱門的研究議題之一。若能準確分類評論文章、網誌內容，將有助於產品或服務上的競爭分析或了解大眾在公共議題上的意見傾向。本論文提出一個基於評論目標發掘及意見詞修飾關係之微網誌評論意見傾向計算方法。首先，從微網誌收集主題相關評論及句子簡化處理。接著根據評論主題以及意見詞的修飾關係，發掘出主題相關的評論目標以判斷其意見傾向。實驗針對 50 部電影在 Twitter 上的 1000 篇英文評論進行分析，結果顯示本論文方法平均準確率 accuracy 為 84.44%，同時最高 precision 可達 88.89%，優於 SVM 及 Naïve Bayes 分類法。由此可驗證意見詞修飾關係的規則判斷能有效提高意見傾向分類的準確率。

Abstract

Opinion analysis has grown to be one of the most active research areas in natural language processing. If we can classify reviews and messages of blogs correctly, it will help to analyze product and service competition and to realize the opinion orientations of the people on public issues. In this paper, we propose an opinion orientation estimation approach based on target finding and opinion modifying relations in microblog reviews. First, it collects reviews from microblog and preprocesses the source data. Then, by extracting any entity or aspect of the entity about which an opinion has been expressed according to opinion modifying relations, we calculate the overall score of opinion orientation.

In our experiment on the 1000 movie reviews of 50 movies from Twitter, the average

accuracy of the proposed method is 84.44%, and the highest precision is 88.89%, which is better than SVM and Naive Bayes. This validates the higher precision from modifying relation identification for opinion orientation classification.

關鍵詞：情感分析，意見傾向，微網誌，意見詞，修飾關係

Keywords: opinion analysis, opinion orientation, microblog, sentiment word, modifying relation

一、緒論

情緒偵測 (emotion detection)的發展對於商業與科技的互動具有高度的應用價值，包括依照使用者情緒推薦相符合的文章、音樂等商品。本研究透過知名微網誌 Twitter 的英文短句中的情緒詞彙進行推文 (tweet) 情緒分類，因為短篇文件所包含的語境和詞彙通常比較不足夠，所以短篇文件的文件分類效果通常會比長篇的文件分類效果不佳。有別於傳統文件分類，我們分析情緒詞彙與修飾關係進行以句子為基礎的情緒偵測 (sentence-based emotion detection) 問題。

本研究方法使用英文文法的修飾關係，主要是 tweet 中的內容評論目標與意見詞之間的修飾關係，找到修飾關係即能判斷評論者藉此內容抒發某種情緒。根據評論主題以及意見詞的修飾關係，發掘出主題相關的評論目標以判斷其意見傾向來預測未知情緒類別的文章之可能情緒。

二、相關研究

常見的情緒偵測方法所適用的範圍可分為為句子層次的推論，段落層次和全篇文章層次的情緒偵測方法 [1][2]。因為微網誌的字數限制，本篇研究專注在句子層次的情緒偵測方法，包括評論目標 (target)、意見詞 (opinion word)等。接下來將探討一些使用文件分類相關技術於情緒偵測的文獻，彼此最大的差異在於偵測方法上的差異。

(一)、評論目標發掘

Kim 和 Hovy[3]針對各類主題的新聞進行找出內文中的 opinion holders 及 opinion topics。首先以動詞及形容詞為主建立情緒辭典，接著然後使用剖析器解析句子，並將 FrameNet 的 frame element 及範例句子來進行 Maximum Entropy 訓練以找到句子中的 opinion holder 及 opinion topic。最後實驗結果的準確度為 64%，說明 FrameNet 中的 frame 及 frame element 有限，只能找到部分的句子結構，因此準確率並不高。

Popescu 和 Etzioni[4]針對商品的使用者評論，採用 PMI (Point-wise Mutual Information) 來獲得與主題共同出現機率最高的詞來當作評論目標 (opinion targets)，與本研究分法類似。我們除了使用 PMI 以外，有鑑於不同使用者所使用的字詞會有所不同，所以也將 PMI 所獲得詞的同義字來擴增我們的評論目標。

Qiu 等人[5]使用消費者評論資料集[20]裡的評論辨識出評論目標(target)及意見詞(opinion word)。首先使用 POS tagging 標註字的詞性，他們定義 target 為名詞，opinion words 為形容詞。再透過 Minipar[21]解析句子的結構。最後利用句子的結構及資料已標記好的目標（商品相關屬性等等）和意見詞來找到句子中未知而可能的目標。例如”Canon G3 has a great lens.”，句子經過解析後得到 G3 → subj, lens → obj, G3 為動詞”has”的主詞，lens 為動詞”has”受詞，若已知的目標為”lens”，透過句子的結構關係與已知的目標得知”lens”與”G3”為相同主題的目標。例如”iPod is the best mp3 player”，句子經過解析後得到 iPod → subj, best 修飾 player，而”best”為已知的意見詞，透過句子的結構與已知的意見詞得知”player”與”iPod”為相同主題的目標。

(二)、修飾關係辨識

Zhuang 等人[6]針對電影評論進行情緒的分類。他們使用 Stanford Parser 工具[22]來解析句子結構並找出字與字的修飾關係，進一步定義意見詞的情緒傾向。因為 tweet 句子結構複雜並包含許多口語化用字，若使用 Stanford Parser 來解析 tweet 並不能準確分析字與字間的修飾關係，因此本研究方法定義意見詞與主題之間修飾關係的方法，來克服 tweet 不規則的句子結構。

Qiu 等人[5]透過資料集[20]中已標註的目標及意見詞以及句子的結構關係，來擴增與主題相關的目标及意見詞。在句子中結構使用語意相依法則 (semantic dependency grammars)，可分為直接相依性 (Direct Dependency, DD) 及非直接相依性 (Indirect Dependency, IDD) 兩種字詞結構[23]。透過剖析器將句子，剖析出樹狀的詞性架構，並標記出中心詞(head)所在的位置。以中心詞為基準，考慮其它詞與中心詞的關係。他們針對結構化且單純的評論文章進行實驗，且每篇文章的評論目標較為明確，因此使用剖析工具來解析句子結構較為適合，且能利用資料集已提供與主題相關的目标及意見詞來進一步擴增與主題相關的目标及意見詞。但社群上的訊息結構複雜且訊息內容無特定主題，因此使用剖析工具無法正確解析句子，且因訊息內容主題不明確，不易利用句子結構及修飾關係來找到其他與主題相關的目标及意見詞。因此本論文提出利用統計方法找到可能與主題相關的目标，我們定義意見詞為形容詞及動詞，增加了修飾規則的判斷。並且利用意見詞及目標的距離來判斷修飾關係的可能性，因此不需要考慮句子複雜的結構問題。

(三)、Twitter 與意見分析

Go 等人[7]對 Twitter 進行情緒分析，利用 SVM、Naïve Bayes 及 Maximum Entropy 等分類方法進行比較，並使用 *n*-grams 當作特徵進行分類器的訓練。他們訓練及測試的資料是根據 tweet 中的表情符號來當作情緒分類的標準(正和負)。Pak 及 Paroubek[8]等人根據形容詞的頻率和 Naïve Bayes Classifier 分類訊息，他們亦是使用訊息中的表情符號當作參考答案。

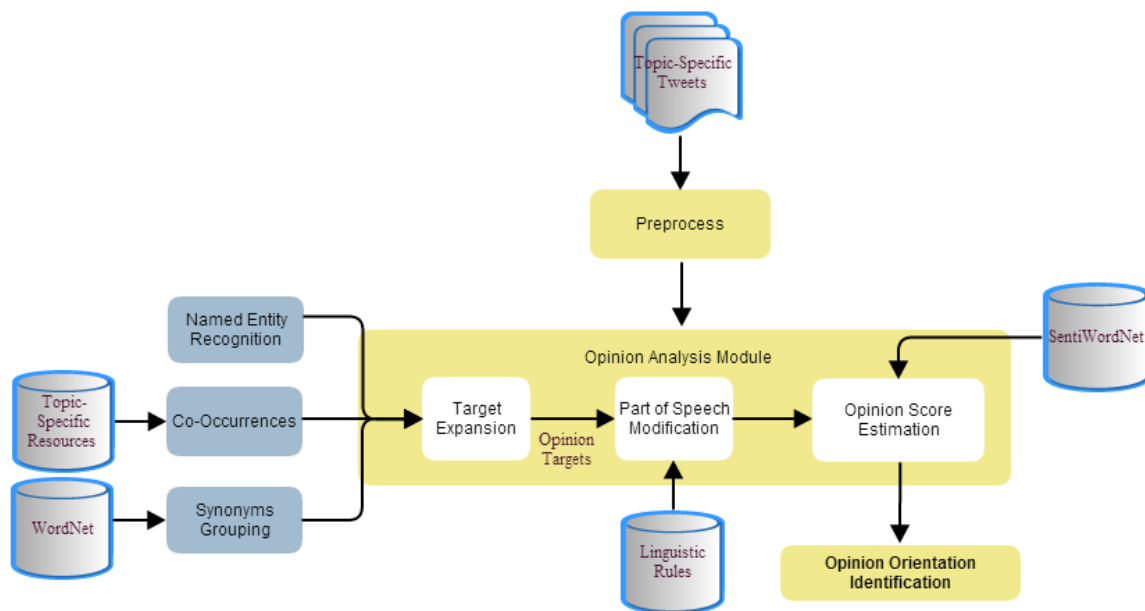
根據以上的研究，因為 Twitter 並沒有提供情緒分類的語料庫，所以往往使用訊息中的表情符號來當作訓練及測試資料的分類標準答案[2][9][10][11]。在本論文中，我們

利用人工標註方法來標記每筆資料的類別 (positive、negative 及 neutral)，如此能針對評論者針對目標 (target)的意見傾向 (opinion orientation)。使用 *n*-grams 及字詞頻率當作訓練特徵常常無法更進一步知道句子結構及語意，本研究方法專注在意見詞及評論目標的修飾關係，能有效知道評論者真正想要評論的事件。

三、研究方法

(一)、系統架構

每則推文 (tweet)的內容可能包含心情、興趣以及對時事的評論等。本論文主要收集使用者在 Twitter 中對主題發表的意見評論，根據意見詞與主題相關目標 (target)之間的字詞修飾關係，精確計算出該主題之總評價。本方法共分為三個主要部分:資料收集、前處理、意見詞修飾關係辨識，架構如圖一所示。



圖一、方法架構圖

如圖一所示，首先透過 Crawler 收集內容包含主題的 tweet (Twitter API[24])，並做前置處理，如：句子簡化，另外根據查詢主題的不同，我們利用主題相關資源 (topic-specific resource)進行目標發掘 (target finding)。以電影為例，我們從全球最大的電影查詢資料庫 IMDB (The Internet Movie Database)中收集電影的相關作者、導演、演員及類型等資訊。接著透過意見分析模組分析每則 tweet 內容是否含有對該主題相關目標的評論，進而計算意見分數來判斷評價的正負面。

(二)、前處理

為了更容易找到訊息中的意見詞 (opinion word)及評論主題 (topic)，我們將與主題相關的 tweet 做前置處理來達到簡化句子的目的。

1、拼字檢查

在 Twitter 中，使用者往往不會太留意拼字的正確性，有時也會藉由單字來強調他欲評論的事件，例如：“I lovvvvvvvvve this movie.”。然而錯誤的拼字可能會影響字詞修飾關係的判斷，因此本論文使用 Google Spell Check[25]對每則 tweet 進行拼字的檢查，在此部份我們是做純拼字錯誤的檢查，對於文法及字義使用不當並無做檢查。

2、Stemming

由於名詞的單複數 (如 movie 和 movies)、詞性的變化 (如 good 和 goodness，動詞的時態 (如 see 和 seeing)，導致語意大致相同的詞或字卻有不同的呈現方式，為了要簡化句子的複雜程度，本研究使用 Porter Stemming algorithm 來進行字根還原的處理。希望將這些後綴去除同時並不影響文字本身的意義，而且對於檢查查全率的提升也更有幫助。

3、特徵過濾

Tweet除了內容本文之外，還包含以下幾點特徵:

- Username: 給一個Tweet的回覆或留言。用法為在 @ 符號後加對方的 Twitter ID，一個空格或冒號後寫上回覆內容。例如，“@disc the tall man is such a good movie.”。
- Links (url): 使用者常會在tweet中分享鏈結，例如：“That Blade Runner sequel is still happening. After seeing Prometheus, I was hoping everyone had forgotten about it. <http://www.deadline.com/hollywood>.”。
- Retweet (RT): 就是轉推的意思，當你在Twitter上看到一個有意思的tweet，就可以RT一下，以幫助傳播這條信息。用法為：RT @原始發布者Twitter ID: 被轉推的原文。例如：“RT if you like Titanic, Harry Potter, Twilight, Pitch Perfect, Skyfall, Life of Pi, Transformers, Les Miserables & etc.. :)”。

以上特徵並不影響使用者在tweet中欲表達的敘述內容，但會使得訊息內容複雜而影響到意見分析的準確率，所以我們將這些特徵予以刪除，只留下敘述內容。

4、詞性標註 (POS Tagging)

因為研究中須找出詞與詞中的修飾及對等關係，任何語言處理的系統都必須先分辨文本中的詞才能進行進一步的處理，我們使用 Stanford POS Tagger[26]進行詞性標註。

(三)、意見分析

本章節說明意見分析的方法，主要分為三個步驟：發掘相關的評論目標 (target expansion)，意見詞與評論目標修飾關係 (opinion words modification relation)，最後計算句子的意見分數 (opinion Score estimation)來判斷句子的意見傾向 (opinion orientation identification)。

1. 評論目標發掘

雖然我們從Twitter中依照主題收集tweet，但tweet內容仍可能包含發文者對無關主題的評論或敘述，例如:I went to theater to watch Argo yesterday. Ring Ring Ring! I was so humiliated when my phone rang out.，例子中屬於負面的情緒抒發，但評論的是因為電影中電話響起而感到丟臉，此內容並沒有針對Argo這部電影做任何的評價及論述。

為了能精確計算出使用者在Twitter中對主題的評價，首先我們要找到tweet中使用者可能在評論的事件，並且要確認此事件是否跟主題相關。例如: “I watched battleship last night, Rihanna’s acting is amazing.” 例子中，在講述battleship這部電影中演員的演技很不錯，由此例可發現，發文者並不直接評論battleship，而是對演員的演技做好的評價，這是一篇對battleship正面評價的tweet，因為演員的名字也是電影的屬性之一。目標(target)為與主題高度相關的字詞，可能是同義詞或評論主題使用的字詞，本小節將介紹我們找target的方法。

(1)、命名實體辨識

在本研究中以電影為例，因為演員、導演、編劇等人物都極可能是評論電影的網友可能評論的目標，專有名詞的標記，可以解決詞庫涵蓋不足的問題，也因其牽涉到人、事、時、地、物等重要內容，我們使用 Stanford Named Entity Recognizer[27]來做專有名詞標記，主要是要找出 tweet 中可能出現跟電影有關的專有名詞，例如：演員、導演、編劇、其他電影專業術語等。

(2)、共同出現關係 (Co-Occurrence)

我們在 Movie Review Data 文集[30]使用 PMI (Point-wise Mutual Information)處理詞彙共同出現關係 (word collocation)，進而了解文集中使用者評論電影時最常提及的名詞。我們利用此文集所有的名詞單獨出現次數 (term frequency, tf) 和與單字”movie” 共同出現的詞彙組合 (emotion-words collocation pairs)出現次數，分別計算 PMI score 並依照降冪排列，再從中取 PMI 最高的前 k 個詞彙共同出現的組合作為特徵子集合。最後在實驗中評估 k 應該取幾個字來當作我們的 target。

$$PMI(w1,w2) = \log_2 \frac{P(w1,w2)}{P(w1)P(w2)} \quad (1)$$

如公式 1 所示，P(w1)和 P(w2)可以透過計算 w1 和 w2 個別出現的次數作為機率估計值；而 P(w1,w2)代表 w1 和 w2 兩個字共同出現 (co-occurrence)的機率，可以透過計算兩個字在文章中共同出現的次數作為機率估計值。

(3)、同義詞

在英文中，同一事物卻有很多單字可以表達，例如：與”movie”同義的單字有”film”、”show”、”flick”、”motion picture”、”moving picture”等等。為了增加與主題相關的 target 數量，我們使用 WordNet 來找出先前找到的 target 的同義字。

2. 詞性修飾關係

在找到可能的評論目標 (target)後，接下來要找到在句子中評論這些目標的修飾關係，並根據情緒詞典比對，計算出對評論目標的情緒分數：

(1)、修飾關係辨識

本方法在搜尋修飾關係時都是以句子為單位，多個單字放在一起可以表示出完整的意思，且通常以標點符號將句子間做區隔。

一個句子的基本結構包含兩個重要的部分：主詞部分 (subject group) 和述語部分 (predicate group)，亦即一個句子必須要有主詞和述語動詞。而意見詞常是動詞以及形容詞[3]，所以本論文僅找句子中動詞及形容詞與目標的修飾關係[12][13]。為了要辨別目標的評論與其他敘述，我們訂定以下修飾規則，若句子中包含以下的修飾關係，則此句子可能含有對目標的評論。我們將意見詞的意見傾向及分數作為此修飾關係的意見傾向及分數：

1. VB/ VBD/ VBG/ VBN/ VBP/ VBZ (意見詞，動詞) + T: T 為 target，也是動詞之後的補語，顧名思義，就是針對動詞，再多作描述，補充動詞不足之處，表達出句子完整的意思，補足方式通常是以名詞或代名詞作為動詞的受詞。例如:I love battleship.，"love"是句子中的 VB，"battleship"是電影名字也是我們的 target，在此句子中就是"love"的受詞，所以符合我們的此項規則;因為"love"是屬於正面情緒詞，所以此句是對電影"Battleship"是屬於正面的評價。
2. T + VB/ VBD/ VBG/ VBN/ VBP/ VBZ (意見詞，動詞): T 為 target，是句子中的主詞。主詞之後若是動詞，那此動詞可能在描述主詞的行為或狀態。例如: The film bored me to death.，"film"是我們找到的 target，"bored"是句子中的 VB，因為"bored"是屬於負面情緒詞，所以此句是屬於對電影的負面的評價。
3. T + VB/ VBD/ VBG/ VBN/ VBP/ VBZ + JJ (意見詞，形容詞): T 為 target，是句子中的主詞。例如: This movie is worth seeing.，"movie"是我們找到的 target，"is"是句子中的 VBZ，"worth"在句子中為 JJ，因為"worth"是屬於正面情緒詞，所以此句是屬於對電影的正面的評價。
4. JJ (意見詞，形容詞) + T: 此項規則主要是找到 target 及修飾 target 的修飾詞。例如:It's my favorite movie.，例子中，"movie"是 target，"favorite"是 JJ，也是修飾 target 的形容詞，因為"favorite"是屬於正面情緒詞，所以此句是屬於對電影的正面的評價。

以上的修飾關係，都是尋找句子中距離最近的單字，例如: In the first movie Tony Curtis's acting is amazing.，例子中找到第 3 種特徵 T + VBZ + JJ，但在 target finding 時我們找到"movie"及"acting"兩個 target，"is"是 VBZ，這時會尋找與修飾詞 JJ "amazing"最近的字，也就是"acting"，最後找到的特徵就是"acting + is + amazing"。在此特徵中會因為字與字在句子中的距離而影響正負面評價的分數，我們將在後面章節作介紹。

(2)、比較句

所謂「比較級」就是在雙方或兩者間做比較的表達方式，比較的內容當然就不外是

「形容詞」或「副詞」了，使用者在評論某一物件時常會以相似的物件來做比較，例如：
The picture quality of Camera-x is better than that of Camera-y.在例子中比較兩台相機的相片畫質，這是最常見的比較關係。我們將常見的比較關係分為下列兩項[14][15]:

1. 非對等比較 (non-equal comparisons)：物件間比較屬性的優劣，例如：**The VIA chip is faster than that of AMD.**，例子中是最常見的比較物件屬性的優劣關係。又例如：**I prefer VIA to AMD.**，此例子也是表達優劣關係。我們利用 POS tagger 來標記比較級，標記為”JJR”、”JJS”、”RBR”、”RBS”，為形容詞及副詞的比較級，例如：**Life was harder then because neither of us had a job.**，例子中”harder”經過 POS tagger 標註為”JJR”。再來尋找 target 與標記的相對位置：
 - Target + JJR / RBR (意見詞): 若 JJR / RBR 是正面情緒詞，則 tweet 對此 Target 是屬於正面評價，若 JJR / RBR 是負面情緒詞，則 tweet 對此 Target 是屬於負面評價。
 - JJR / RBR(意見詞) + than + Target：若 JJR / RBR 是正面情緒詞，則 tweet 對此 Target 是屬於負面評價，若 JJR / RBR 是負面情緒詞，則 tweet 對此 Target 是屬於正面評價。例如：**Why are books always better than the movie versions?**，例子中，若 movie 為 target，句中找到”better than”為形容詞比較級，屬於正面評價，但在此比較關係中，”book”優於”movie”，所以對於”movie”是屬於負面的評價
2. 對等比較 (equal comparisons)：比較的關係的程度或強弱是相等的，例如：**The picture quality of Camera-x is as good as that of Camera-y.**，例子中是說明 Camera-x 的相片品質與 Camera-y 一樣好。此用法是英文文法中的常用特定模式。若 tweet 內容找到”as + 原級形容詞 + as”規則，若此原級形容詞為正面情緒詞，則此 tweet 對此 target 有可能是正面評價;反之，若此原級形容詞為負面情緒詞，則此 tweet 對此 target 有可能是負面評價。

(3)、否定詞

根據 Tottie[16]，英文的否定標記 (negative marker)主要分為三大類：

- (1) not 否定 (not-negation)
- (2) no 否定 (no-negation)
- (3) 詞綴否定 (affixal negation)

否定標記的範例如表一所示：

表一、否定標記

| not-negation | no-negation | affixal negation |
|--------------|--------------------------|------------------|
| not | No | (im)perfect |
| | nor | (ir)respective |
| | none | (in)dependent |
| | never | (un)able |
| | neither | (non)functional |
| | nowhere, nothing, nobody | meaning(less) |

由 Tottie 的定義中可以發現，not 否定和 no 否定基本上屬於語法的範疇，而詞綴否定（affixal negation）則是在詞彙的範疇。在詞綴否定的部分在 SentiWordNet 能作適當的辨別，例如：perfect → positive，imperfect → negative。

not-negation 與 no-negation 的處理[17]，先依照先前介紹的方法找到修飾關係及正負面情緒，若是句子中含有 not-negation 與 no-negation 的字，則會反轉正負面結果，例如：I don't like this movie, the plot is so boring. 例子中，依照之前介紹的規則在” I don't like this movie,”句子中的”like this movie”找到 VB + Target 的修飾關係，屬於正面評價的句子，但在句中找到”n't”的否定詞，所以原屬正面評價的句子在最後會反轉成負面評價。

3. 意見評分

Tweets 在經過前章節的修飾關係特徵的搜尋，這些修飾關係會因為字與字之間的距離而影響到情緒分數，例如：In the first movie Tony Curtis's acting is amazing.，例子中，找到”T + VBZ + JJ”的修飾關係”acting + is + amazing”，我們會計算意見詞與 target 之間的距離來調整修飾的權重。一則 tweet 中可能存在許多修飾關係的特徵，所以需要經過正負面情緒分數的加總來判斷此 tweet 是屬於正面情緒或負面情緒，亦或是客觀論述的 tweet。針對某句子 s，其情緒分數的計算如下：

$$\text{score}(s) = \sum_{op_j \in s} \frac{op_j \cdot so}{d(op_j, t_i)}, \quad t_i \in T \quad (2)$$

公式中， op_j 是句子 s 中的意見詞，T 為經由 target finding 所找到 target 的集合， $d(op_j, t_i)$ 是在句子 s 中意見詞 op_j 及 t_i 的距離，so 是修飾字 op_j 的情緒面向分數，由 SentiWordNet 得知。公式的 multiplicative inverse 是為了判斷修飾字在 f_i 修飾 target 的可能性，若距離越遠則計算出的情緒分數越低。整篇 tweet 的分數即為所有句子情緒分數的總和。最後依據分數將 tweet 分成三類：

- 正面(positive): score > 0。
- 負面(negative): score < 0。
- 客觀(objectivity): score = 0。

四、實驗與討論

(一)、測試資料收集

隨機挑選在 2013 年 2 月至 3 月上映的五十部電影收集其相關評論 tweet。因為不想使資料過於集中在某一天，所以每隔 5 天收集一次。收集日期分別為 2013/3/20、2013/3/25、2013/3/30、2013/4/5 及 2013/4/10，每日的收集量為 200 則 tweets，測試資料總共 1000 則 tweets，接著使用人工標註每則 tweet 的情緒面向作為實驗的標準答案，由 5 人進行情緒標註，標註有三類：正面、負面、客觀。若是正面為+1，負面為-1，客觀為 0。最後依三種分數的個別加總，採多數決的方式來決定每則 tweet 的情緒面向。

(二)、實驗結果與討論

我們在進行研究方法中的 target expansion 的實驗與討論，最後比較本論文的方法與 SVM 及 Naive Bayes 分類方法的效果。分別會計算出正負面及主觀(subjectivity)評論的精確率(precision)、查全率(recall)、F1 及準確率(accuracy)等數值來衡量方法效果。

其中 baseline 為未經過評論目標發掘，所以 target 只包含電影名字，藉此來比較評論目標發掘方法的效果。

1. 命名實體辨識

Baseline 因為 target 過於稀少，使得 recall 都過於偏低，無法有效的辨識大部分有關電影的評論。電影評論的 target 可能也包含人名及專有名詞，我們使用 Stanford Named Entity Recognition 工具擴增評論目標的數量。加入 NER 前後的分類效果如表二、表三所示。

表二、加入 Named Entity Recognition 前後的主客觀評論分類效果比較

| 主觀分類 | Baseline | with Named Entity Recognition | Improvement (%) |
|-----------|----------|-------------------------------|-----------------|
| Recall | 0.38173 | 0.40315 | 5.6% |
| Precision | 0.89247 | 0.90671 | 1.6% |
| F score | 0.53474 | 0.55814 | 4.4% |
| Accuracy | 0.57438 | 0.59594 | 3.8% |

雖然使用 Named Entity Recognition 會增加電影相關的導演、演員等的名字及專有名詞，但也會使與電影無關的名字及專有名詞也會納入 target，但因為我們根據電影名稱去收集 tweet，所以大部份的 tweet 內容都是在評論電影，如表二所示，在判斷主客觀評論會因為 target 的增加在 precision、recall 及 accuracy 都有提升。

表三、加入 Named Entity Recognition 前後的正負面評論分類效果比較

| 正負面評論 | Baseline | with Named Entity Recognition | Improvement (%) |
|--------------------|----------|-------------------------------|-----------------|
| Positive Recall | 0.402 | 0.41673 | 3.7% |
| Negative Recall | 0.39483 | 0.41837 | 5.9% |
| Positive Precision | 0.91224 | 0.92194 | 1.1% |
| Negative Precision | 0.88563 | 0.90173 | 1.8% |
| Positive F score | 0.55807 | 0.57400 | 2.9% |
| Negative F score | 0.54616 | 0.57156 | 4.7% |
| Accuracy | 0.58813 | 0.60956 | 3.6% |

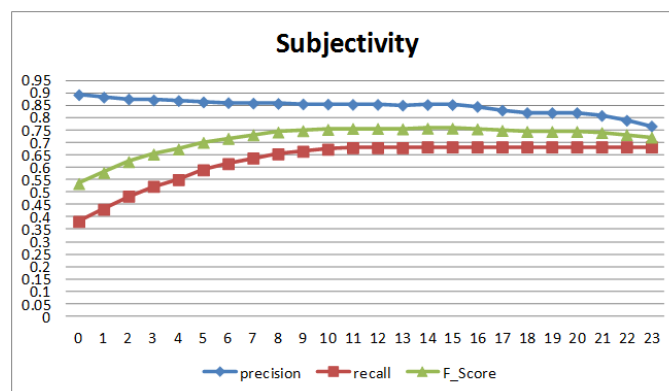
如表三所示，隨著主客觀評論分類的效果提升，進一步將主觀評論分辨正面(positive)及負面(negative)的效能也能有所提升。

如表二及表三所示，tweet 中所提及的名字及專有名詞不一定都和電影相關，所以

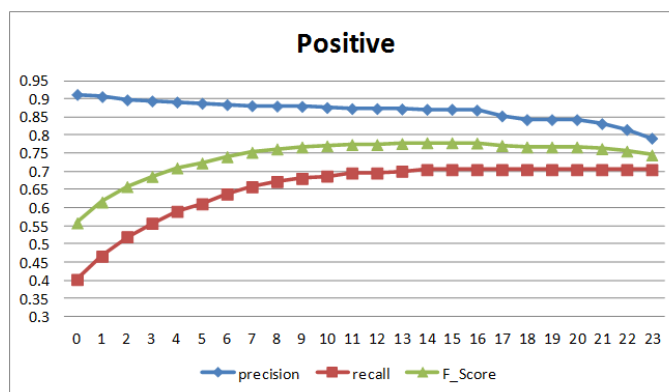
NER 在擷取人名及專有名詞時可能會找到無關的 target，因此在 precision 只有稍微進步。

2. 共同出現關係

如圖二、圖三所示，我們分別觀察主客觀及正面情緒評論的分類效果，當 k 值由 0 到 15 時，因為增加 target 的數量，recall 是持續且明顯的上升，precision 可能因為當考慮的 opinion target 增加使得修飾關係變得複雜而呈現持續下降的趨勢，但 accuracy 的值也因為 k 值的增加而有明顯上升。當 k 值由 16 到 23 時，target 與電影的相關度下降，precision 仍然是呈現持續下降的趨勢，而 recall 並沒有再上升，所以 F1 score 仍然在持續下降。accuracy 值也因為預測失敗逐漸下降。



圖二、不同 k 值 PMI 對主客觀評論分類的 precision、recall 及 accuracy 影響



圖三、不同 k 值 PMI 對正面情緒分類的 precision、recall 及 accuracy 影響

根據表四、表五我們觀察到 precision 是下降的，當參考越多的 target，就會增加句子分析的複雜度。較簡單的句子能夠判斷正確例如：“Loving the **music** in **total recall** :-)” ，例句是在評論“Total Recall”這部電影的音樂。例子中找到意見詞“loving”修飾 target “music”。複雜的句子，例如：“Watching **Warm Bodies!** :D right after i listen to my 5SOS playlist.... I have a problem... Im addicted to 5SOS **music**...” ，例句中出現 “Warm Bodies”及“music”這兩個 target，找到兩組修飾關係：一為 Watching 修飾 Warm Bodies，系統判斷為 objective，另一為 addicted 修飾 music，系統判斷為 positive。最後經過分數的加總判斷此 tweet 為 positive。然而此評論並不是針對電影的音樂，所以此 tweet 應為“objective”。雖然有找到電影名字的修飾關係，但 target 並不一定與該部電影相關。

表四、加入 PMI ($k=15$)前後的主客觀評論分類效果比較

| 主觀分類 | Baseline | $k = 15$ | Improvement (%) |
|-----------|----------|----------|-----------------|
| Recall | 0.38173 | 0.68176 | 78.6% |
| Precision | 0.89247 | 0.85247 | -4.4% |
| F score | 0.53474 | 0.75761 | 41.7% |
| Accuracy | 0.57438 | 0.79341 | 38.1% |

表五、加入 PMI ($k=15$)前後的正負面評論分類效果比較

| 正負面評論 | Baseline | $k = 15$ | Improvement (%) |
|--------------------|----------|----------|-----------------|
| Positive Recall | 0.402 | 0.70449 | 75.2% |
| Negative Recall | 0.39483 | 0.67968 | 72.1% |
| Positive Precision | 0.91224 | 0.87124 | -4.5% |
| Negative Precision | 0.88563 | 0.84963 | -4.1% |
| Positive F score | 0.55807 | 0.77904 | 39.6% |
| Negative F score | 0.54616 | 0.75521 | 38.3% |
| Accuracy | 0.58813 | 0.82732 | 40.7% |

3. 同義詞

如表六所示，經由同義詞來擴增 target 的數量，在主客觀評論分類的實驗數值都有所提升，原因是能克服不同使用者評論同一事物卻有很多單字可以表達的問題，因此在 tweet 中能找到更多與電影有關的評論及修飾關係。

表六、加入同義詞前後之主客觀評論分類效果比較

| 主觀分類 | PMI ($k = 15$) | PMI ($k = 15$) + synonyms | Improvement (%) |
|----------|------------------|-----------------------------|-----------------|
| F score | 0.75761 | 0.77374 | 2.1 % |
| Accuracy | 0.79341 | 0.80971 | 2.0 % |

如表七所示，因為判斷主觀評論的效果增加，在正負面評論分類效能也有所提升。這說明同一件事物或事件，不同使用者會使用不同字眼來評論或敘述，所以經由找同義字能彌補這方面的不足。

表七、同義詞前後之正負面評論分類效果比較

| 正負面評論 | PMI ($k = 15$) | PM I($k = 15$) + synonyms | Improvement (%) |
|------------------|------------------|-----------------------------|-----------------|
| Positive F score | 0.77904 | 0.79307 | 1.8 % |
| Negative F score | 0.75521 | 0.76201 | 0.9 % |
| Accuracy | 0.82732 | 0.83929 | 1.4 % |

4. 所提方法與 SVM 及 Naïve Bayes Classifier 之比較與討論

我們利用 n-gram 將訊息內容作切割，所找到的分割字當作一組獨立的詞彙，因為

訊息內容較短，我們使用 unigram 和 bigram。訊息透過 n-gram 斷詞演算法得出的 n-gram 特徵值套入支持向量機(Support Vector Machine, SVM)及貝氏分類器(Naive Bayes Classifier)來分析詞頻進行自動訊息分類[18]。最後比較本研究方法以及傳統分類器使用字的特徵進行短訊息的情緒分類的效果。

透過前面實驗的觀察，發現在 target expansion 使用 Named Entity Recognizer、PMI ($k = 15$)及 Synonyms 可提升分類的效果，因此，這裡我們在 target expansion 使用上述方法並透過修飾關係的分數計算進行實驗，並與 SVM 及 Naive Bayes 分類的效果比較。

從表八得知，我們的分法無論在 positive、negative 及 subjectivity 的精確率明顯優於 SVM 及 Naive Bayes。

表八、所提方法與 LibSVM 及 Naive Bayes Classifier precision 的比較

| | 所提方法 | LibSVM | Naive Bayes |
|------------------------|---------|---------|-------------|
| Positive Precision | 0.88893 | 0.72810 | 0.68017 |
| Negative Precision | 0.85392 | 0.69724 | 0.63694 |
| Subjectivity Precision | 0.87269 | 0.69378 | 0.63954 |

因為本研究是針對評論目標與意見詞之間的修飾關係，所以只有在句子中找的特徵才會判斷情緒面向，可能的目標較侷限，無法找出所有評論者可能評論的目標，所以在 recall 的效果會低於 SVM 及 Naive Bayes，如表九所示。

表九、所提方法與 LibSVM 及 Naive Bayes Classifier recall 的比較

| | 所提方法 | LibSVM | Naive Bayes |
|---------------------|---------|---------|-------------|
| Positive Recall | 0.72718 | 0.88531 | 0.92423 |
| Negative Recall | 0.69796 | 0.87938 | 0.77431 |
| Subjectivity Recall | 0.71284 | 0.88157 | 0.90767 |

從表十得知，本研究方法的 accuracy 優於 SVM，是因為有較好的精確率。根據實驗中的 precision 及 accuracy，本研究能夠針對短訊息中的某特定主題做有效主、客觀評論的分類，並且能進一步將主觀評論精確地分類出正、負面的情緒傾向。

表十、所提方法與 LibSVM 及 Naive Bayes Classifier accuracy 的比較

| | 所提方法 | LibSVM | Naive Bayes |
|----------------|---------|---------|-------------|
| 主觀分類 Accuracy | 0.82271 | 0.81673 | 0.78923 |
| 正負面評論 Accuracy | 0.84439 | 0.83439 | 0.81195 |

五、結論

本論文提出一個基於評論目標發掘及意見詞修飾關係之微網誌評論內容意見傾向計算方法。根據評論主題以及意見詞的修飾關係，發掘出主題相關的評論目標以判斷其意見傾向。然而若要提高分類準確率，還需要進一步找出 opinion holder 與 opinion polarity，甚至在時間軸上的變化關係，是屬於比較高階的應用。這時候文句若能先標詞性、標片語、甚至到句型剖析等前處理，並擷取與主題相關的屬性來增加找出 opinion holder 的準確度，以上都對提高情緒分析的準確率會有幫助。

Twitter 在發文的數字上有限制，主要是以平常口語、簡短的方式在 Twitter 上發佈，因此常有俚語的部分，所以在字串處理方面會有困難。例如：“shh identity thief, is it good movie I gonna bring my Lil g Cuzco to DE movies, Rollin out Tass can't wait till class finish lol.”，出現很多簡寫及口語化的字，在前處理時的字串處理造成困難，以至於影響詞性標註、評論目標發掘以及修飾關係，所以會降低本方法的準確率。目前也沒有較正式的 tweet 電影評論資料集，本實驗是自己收集資料集並且利用人工標註方法來做資料集的情緒標註，在數量上較為不足。以上的難題都是未來待克服的議題。

參考文獻

- [1] Devillers L., Vasilescu I., and Lamel L., “Annotation and Detection of Emotion in a Task Oriented Human-Human Dialog Corpus,” Proceedings of the ISLE Workshop on Dialogue Tagging for Multi-Modal Human-Computer Interaction, pp.624-629, 2002.
- [2] Ang J. C., “Prosodic Cues for Emotion Recognition in Communicator Dialogs,” M.S. thesis, University of California Berkeley, 2002.
- [3] S. Kim and E. Hovy. “Extracting opinions, opinion holders, and topics expressed in online news media text,” In Proceedings of ACL/COLING Workshop on Sentiment and Subjectivity in Text, pp.1-8 ,2006.
- [4] Popescu A. M., Etzioni O., ”Extracting product features and opinions from reviews,” Proceedings of the conference on Human Language Technology and Empirical Methods in Natural Language Processing, p.339-346, October 06-08, Vancouver, British Columbia, Canada, 2005.
- [5] Qiu, G., Liu B., Bu J., and Chen C.. “Opinion word expansion and target extraction through double propagation,” Computational Linguistics, pp.363-370, 2011.
- [6] Zhuang L., Jing F., Zhu X.-Y., “Movie review mining and summarization,” Proceedings of the 15th ACM international conference on Information and knowledge management, Arlington, Virginia, USA, pp. 43-50, November 06-11, 2006
- [7] Go A., Bhayani R., and Huang L., “Twitter Sentiment Classification using Distant Supervision,” Technical report, Stanford Digital Library Technologies Project.
- [8] Pak A. and Paroubek P., ”Twitter as a corpus for sentiment analysis and opinion mining,” Proceedings of LREC, pp.1320-1326, 2010.
- [9] Sun Y. T., Chen C. L., Liu C. C., Liu C. L., Soo V. W., “Sentiment Classification of Short Chinese Sentences,” ROCLING 2010.

- [10]Tang, Y., Chen, H. “Emotion Modeling from Writer/Reader Perspective Using a Microblog Dataset,” In: Workshop on Sentiment Analysis where AI meets Psychology (SAAIP 2011), pp. 11–19. ACL, 2011.
- [11]Davidov, D.; Tsur, O.; and Rappoport, A. “Enhanced sentiment learning using twitter hashtags and smileys,” In Proceedings of Coling, pp.241-249 ,2010.
- [12]Wiebe J., Riloff E.. “Creating subjective and objective sentence classifiers from unannotated texts,” pp.486-497 , In CICLing-2005.
- [13]Barbosa, L., Feng, J. “Robust sentiment detection on twitter from biased and noisy data,” In Proc. of Coling, pp.36-44, 2010.
- [14]Jindal, N., and Liu, B. “Mining comparative sentences and relations,” AAAI’06, 2006.
- [15]Jindal, N. and Liu, B. “Identifying comparative sentences in text documents,” SIGIR-06, 2006.
- [16]Tottie. Gunnel. “Negation in English Speech and Writing: A Study in Variation,” San Diego: Academic Press, 1991.
- [17]H. Zeijlstra. “Negation in Natural Language: On the Form and Meaning of Negative Elements,” Language and Linguistics Compass, 1(5):498—518, 2007.
- [18]B. Pang and L. Lee. “Opinion mining and sentiment analysis,“ Foundations and Trends in Information Retrieval, 2008.
- [19]Tom M. Mitchell, Machine Learning, WCB-McGraw-Hill, 1997
- [20]Customer Review Collection, <http://www.cs.uic.edu/liub/FBS/sentiment-analysis.html>
- [21]Minipar, <http://webdocs.cs.ualberta.ca/~lindek/minipar.htm>
- [22]The Stanford Parser: A statistical parser, <http://nlp.stanford.edu/software/lex-parser.shtml>
- [23]Dependency grammar, http://en.wikipedia.org/wiki/Dependency_grammar
- [24]Twitter API, <https://dev.twitter.com/docs/api>
- [25]Google Spell Checker, <https://code.google.com/p/google-api-spelling-java/>
- [26]Stanford POS Tagger, <http://nlp.stanford.edu/software/tagger.shtml>
- [27]Stanford Named Entity Recognizer, <http://nlp.stanford.edu/software/CRF-NER.shtml>
- [28]Movie Review Data, <http://www.cs.cornell.edu/people/pabo/movie-review-data/>