

情感分析於投資溫度評分之應用

Sentiment Analysis for Investment Atmosphere Scoring

彭志翔 Chih-Hsiang Peng

元智大學資訊管理學系

Department of Information Management College of Informatics

Yuan Ze University

metgnr89@apb.gov.tw

禹良治 Liang-Chih Yu

元智大學資訊管理學系

Department of Information Management College of Informatics

Yuan Ze University

lcyu@saturn.yzu.edu.tw

摘要

網路財經文本和金融商品投資收益的相關性在近年的研究議題上十分熱門。在此類型研究中，以使用自然語言處理領域中的情感分析技術量化投資大眾心理最受研究者青睞。在情感分析中，有類別型及維度型的分析方法，而維度型的分析方法又可分為單維度型與多維度型。本研究使用類別型、單維度型及多維度型三種情感分析方法，將蒐集的網路財經文本量化為投資溫度評分。實驗結果顯示，以多維度型的投資溫度評分與市場漲跌趨勢最具相關性。

Abstract

The correlation between Internet financial texts and financial commodity investment income has been an emerging research topic in recent years. Sentiment analysis is a useful technique to accomplish the investment psychology. In sentiment analysis, there are two main approaches to emotion state representation: categorical and dimensional. The dimensional approach can be further divided into single dimension and multiple dimensions. This study investigates the use of categorical, single-dimensional and multi-dimensional approaches to quantifying the Internet financial texts as investment atmosphere scoring. Experimental results show that the multi-dimensional method achieves the best correlation between the

investment atmosphere scoring and stock price trends.

關鍵詞：自然語言處理，情感分析，相關係數

Keywords: Natural language processing、Sentiment analysis、Correlation coefficient

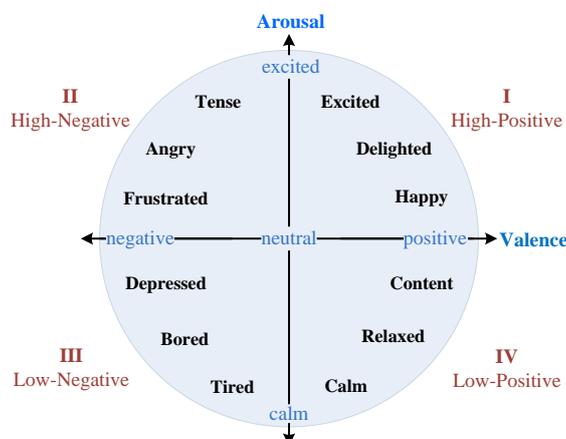
一、緒論

現今民眾從事工商經濟活動後，常會將剩餘可支配所得用來進行投資理財。目前常見的投資理財工具包括了股票、基金及其他衍生性金融商品，在這個高度專業化的社會環境中，民眾無法花太多時間研究、分析產業或區域的發展，也難以掌握精深的金融專業知識。據研究指出，我國投資人透過「網路」獲取投資相關資訊的比例最高，占 80.5%，其次是「報紙雜誌」，占 46.3%[1]。

在投資實務界中，有德國證券界教父之稱的安德烈·科斯托蘭尼（André Kostolany）強調「行情=資金+心理」[2]，認為單靠資金並不能影響股市，還要看投資大眾的心態，也就是「心理因素」，此一論點目前仍被許多投資人奉為圭臬。而現在最能影響投資大眾心態的資訊來源，就是網路。網路時代的投資市場資訊流動快速，但一般人通常沒有能力及時間，即時且完整地掌握網路上大量投資訊息，因此使用新聞報導預測股價走勢一直引起許多研究人員的興趣，此類研究大多係透過文字探勘技術(Text Mining)以及自然語言處理(Natural Language Processing) 中的情感分析(sentiment analysis) [3], [4], [5] 技術達成。

情感分析可分為類別型及維度型 2 種，類別型是將情感區分為正向(positive)、負向(negative)2 類，或是加上中立(neutral)分為 3 類。而維度型情感分析，可分為單維度與多維度 2 種不同的分析方法，單維度指的是採取一個情感程度值來分析，以連續型類值來表示情感的正負向。而多維度是用多個情感特徵來分析，Russell[6]認為情感間存在相關，相關性能夠透過一個 Valence-Arousal(VA)二維空間模型來表示，如圖一。Valence 代表正面或負面的程度，Arousal 則代表激動及冷靜的程度。情感經分析後，可在該二維空間模型上表示為一點。多維度的情感分析運用在財經文本分析上面，例如「美股上漲」與「美股狂漲」都是正向形容美股，但前者的強度並不及後者，影響程度應會有差距。相較於單維度情感分析，多維度情感分析應該更能精確的表達出文本中高度正面或高度負面的情緒。

本研究的動機就是希望比較上述三種情感分析方法將網路財經文本量化的結果，目的是運用情感分析技術將網路財經文本情感量化成投資溫度評分，並以實驗方式驗證投資溫度評分與選定市場趨勢間的相關性，因此本研究將針對以下 3 項進行實驗與探討：



圖一、Valence-Arousal 二維情感模型示意圖[7]

- 1、在網路財經文本的來源上，本研究係使用網路財經新聞，時間區間為 2017 年 1 月至 2020 年 5 月。
- 2、在與投資溫度評分進行相關性檢定的市場選擇上，本研究選擇台股、美股、大中華、亞太市場、新興市場、歐洲等 6 個市場區域。不同於以往多數研究進行個股與網路財經文本情感分析的實驗，本研究認為個股或其新聞較容易受到少數關鍵人士的操控或影響，而國家及區域指數較難被少數人操控，故選擇較大區域的市場趨勢做為檢定目標。
- 3、在投資溫度評分計算上，本研究將計算方式分為類別型、單維度及多維度 3 種計算方式，以實驗驗證哪種計算方式將財經文本情感分析結果量化為投資溫度評分後，與目標市場趨勢最具相關性。

二、相關研究

如依金融學中的效率市場理論，財經新聞出現之際，股價早已反應完畢，因此該理論認為，看新聞做股票係沒有辦法獲利和預測股價，但有許多研究證實投資人的心理因素在大大程度上影響了投資決策，因此投資大眾心理應該是影響金融市場波動不可或缺的一部分。網路財經文本和金融商品投資收益的相關性在近年的研究議題上十分熱門，近年的研究中，在財經文本的來源上，有網路新聞如蔡宇祥[8]的研究係使用微博財經貼文、

Qu et al.的研究係使用 Yahoo Financial News[9]，網路論壇如盧奕叡的研究係使用批踢踢實業坊的文本[10]，社群媒體貼文、搜尋首頁的關鍵字如 Zhao et al.的研究[11]係使用百度搜尋熱門關鍵字。在探討相關性或預測的對象的選擇上，大部分文獻是以個股為主如 Hwang et al.[12]的研究係使用特斯拉股價，部分文獻係採單一市場指數，如，Bourezk et al.[13]的研究係使用摩洛哥指數。文本的情感分析(Sentiment Analysis)，又稱意見挖掘或意見探勘(Opinion Mining)，是運用自然語言處理(Natural Language Processing)、文字探勘(Text Mining)及計算機語言學等技術，進而提取並辨識原文本中作者的情感及主觀意見。常見的情感分析所適用的範圍可分為句子層次、段落層次和文章層次的情感分析方法。類別型的情感分析方法相關研究，Hwang et al.[12]，通過將文本中的正向詞和負向詞計數，以「正詞數」-「負詞數」來計算情感得分，如果情感分數大於 0，則文本標記為「正」，如果小於 0，則文本標記為「負」。維度型情感分析在財經領域之應用並不多見，但已有維度型詞典，如：Affective Norms for English Words (ANEW)[14]、Extended ANEW[15]及語料庫，如：Affective Norms for English Texts (ANET) [16]、EmoBank[17]，相關情感詞向量[18], [19]及迴歸模型[20], [21], [22], [23], [24], [25]也有相關研究發表，均可做為財經領域維度型情感分析之研究材料。

三、研究方法

(一) 中文情感字典

本研究使用中文情感字典 CVAW (The Chinese Valence-Arousal Words) 進行情感分析[16]，CVAW 是一個包含 5,512 個情感多維度型字典，每個詞彙皆包含一組二維的 Valence-Arousal 實數值，維度 Valence 的範圍為 1 至 9，值 1 和 9 分別表示最負面和最正面的情緒表現，值 5 表示沒有特定傾向的中性情緒表現。維度 Arousal 的範圍為 1 至 9，值 1 和 9 分別表示平靜或激動。使用 CVAW 進行預測的實驗結果顯示，與使用英文文本情感分析獲得的結果相當[26]。

(二) 投資溫度

本研究將「投資溫度」定義為網路財經文本的情感，可作為投資大眾心態的參考。

1、財經文本資料蒐集

本研究先蒐集財經新聞做為情感分析之資料，資料來自蘋果日報、蘋果即時、工商時報、

工商即時、中國時報、中時即時、自由即時、中央社、經濟日報、聯合即時，再透過市場區域關鍵字，如：台股、美股、那斯達克、道瓊、標普、陸股、港股、歐股等，將財經文本歸類至特定市場。

依市場區域分類的財經文本，再利用情感字典計算每日的 Valence 及 Arousal 數值，計算結果如圖二所示，其中欄位「V」代表每日所有財經文本 Valence 的平均值，欄位「A」代表每日所有財經文本 Arousal 的平均值，欄位「Pos_num」代表每日所有正向財經文本數量(Valence 值大於或等於 5.5 之文章分類為正向)，欄位「Neg_num」代表每日所有負向財經文本數量(Valence 值小於或等於 4.5 之文章分類為負向)，中性的文章在本研究中不列入計算，欄位「Pos_V」代表每日所有正向財經文本 Valence 的平均值，欄位「Neg_V」代表每日所有負向財經文本 Valence 的平均值，欄位「Pos_A」代表每日所有正向財經文本 Arousal 的平均值，欄位「Neg_A」代表每日所有負向財經文本 Arousal 的平均值，欄位「Pos_VA」是每日所有正向財經文本 Valence 值乘以 Arousal 值之平均值，欄位「Neg_VA」是每日所有負向財經文本 Valence 值乘以 Arousal 值之平均值。

Date	V	A	Pos_num	Neg_num	Pos_V	Pos_A	Neg_V	Neg_A	Pos_VA	Neg_VA
20170101	5.320	5.820	3	2	6.600	5.833	3.400	5.800	27.200	-18.640
20170102	6.200	4.300	2	0	6.200	4.300	0.000	0.000	10.440	0.000
20170103	5.626	4.485	20	4	6.145	4.600	3.600	4.775	106.475	-27.020
20170104	5.721	4.629	11	2	6.255	5.055	3.575	3.550	71.610	-10.905
20170105	5.806	4.727	17	2	6.232	5.059	3.600	4.100	110.023	-11.880
20170106	5.863	4.315	20	0	6.002	4.583	0.000	0.000	92.458	0.000
20170107	None	None	0	0	0.000	0.000	0.000	0.000	0.000	0.000

圖二、財經文本每日 Valence 及 Arousal 數值示例

使用 CVAW 的情感分析結果，以「台股狂瀉近 700 點」為例，Valence 值為 1.5，Arousal 值為 8.5，呈現激動負面的情緒。而「台股重新衝牛市」Valence 值為 9，Arousal 值為 8.3，呈現激動正面的情緒。

2、投資溫度評分計算

本研究將投資溫度評分計算分為類別型、單維度型及多維度型三種。分數計算的時間皆以「週」為單位，較不受短期間過度反應的情緒影響，亦可減少網路財經文本日資料缺失值對實驗結果的影響。

(1)類別型：本研究定義之類別型投資溫度，只將情感區分為正向及負向，不考慮正向及負向的程度問題。計算方式係將一週的正向文章數總和除以一週所有文章數總和，計算公式如下：

$$\sum_{i=1}^n \frac{Pos_num_i}{Pos_num_i + Neg_num_i} \quad (1)$$

(2)單維度型：本研究定義之單維度型投資溫度，是以 Valence 值做為情感區分的依據，Valence 值係介於 1 至 9 之間的連續型數值，數值越接近 1 代表情感越負向，數值越接近 9 代表情感越正向。計算方式係一週內所有正向文章 Valence 值的總和加上一週內所有負向文章 Valence 值的總和，除以一週內所有文章數，再將結果標準化為 0 至 1 的數值，計算公式如下：

$$\frac{\left(\sum_{i=1}^n \frac{Pos_num_i \times Pos_V_i + Neg_num_i \times Neg_V_i}{Pos_num_i + Neg_num_i}\right) - 1}{8} \quad (2)$$

(3)多維度型：本研究定義之多維度型投資溫度，除了考量代表情感正負向的 Valence 值外，還將代表情感激動程度 Arousal 值加入，Arousal 值係介於 1 至 9 之間的連續型數值，數值越接近 1 代表情感越平靜，數值越接近 9 代表情感越激動。計算方式係一週內每日所有正向文章的 Valence 值乘以 Arousal 值相加後，加上一週內每日所有負向文章的 Valence 值乘以 Arousal 值，再將結果標準化為 0 至 1 的數值。計算公式如下：

$$\chi = \sum_{i=1}^n Pos_VA_i + Neg_VA_i \quad (3)$$

$$\chi_{norm} = \frac{\chi - \chi_{min}}{\chi_{max} - \chi_{min}} \in [0,1] \quad (4)$$

3、目標市場參考指數

本研究選定 6 個市場區域各 2 種參考指數做為實驗資料，分別為臺灣(台股加權指數、MSCI Taiwan Index)、美國(S&P500、Nasdaq)、大中華區域(FTSE Greater Chinese Index、MSCI Gold Dragon Index)、亞洲太平洋區域(MSCI Asia Pacific Index、FTSE Asia Pacific Index)、新興市場(MSCI Emerging Markets Index、FTSE Emerging Index)及歐洲(MSCI Europe Index、FTSE Europe Index)，參考指數主要以 MSCI 及 FTSE 所發行的指數為主。參考指數的歷史資料係由 Yahoo Finance 及 Investing.com 取得。

四、實驗結果

(一)實驗流程

在投資溫度評分計算方式部分，本研究使用 3 種方式，分別為類別型、單維度型、多維度型。在市場區域部分本研究選擇 6 個區域，在各區域中分別選定 2 種參考指數，總計 12 種參考指數，在參考指數漲跌趨勢計算部分為本週收盤價減上週收盤價。

1、實驗資料

本研究使用之網路財經文本及參考指數，蒐集時間為 2017 年 1 月 1 日至 2020 年 5 月 31 日。每篇文本經計算完 Valence 值及 Arousal 值後，先將 Valence 值大於或等於 5.5 之文章分類為正向，小於或等於 4.5 之文章分類為負向，介於 4.5 至 5.5 之文章分類為中性文章，本研究將中性文章排除後，共計收集 13 萬 2,564 篇網路財經文本。

2、評估標準

本研究採用皮爾森相關係數(r)做為相關性評估標準，其值在 0.1 以下屬於微弱或無相關，在 0.1 至 0.39 間屬於低度相關，在 0.4 至 0.69 間屬中度相關，0.7 至 0.99 間屬高度相關。

(二)實驗結果與分析

1、實驗結果

投資溫度評分與市場區域參考指數漲跌趨勢的相關性檢定結果如表一所示。

(1)維度型投資溫度評分計算方式結果比類別型為好。

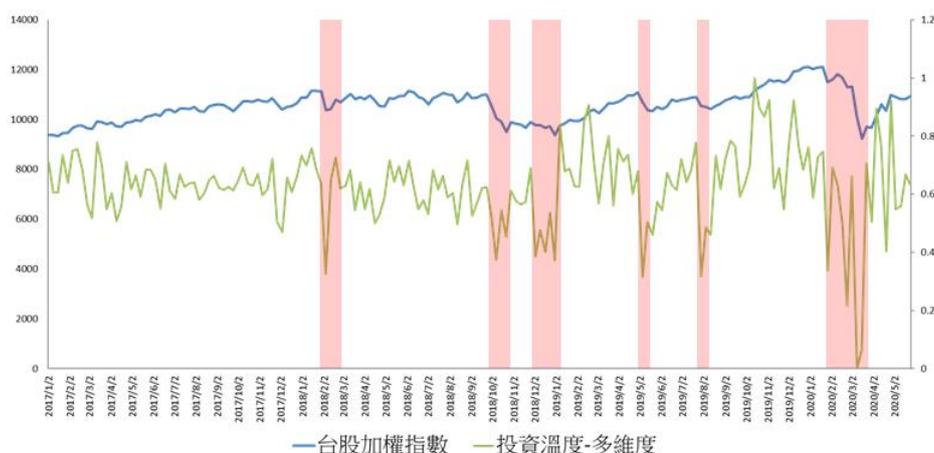
(2)在維度型的投資溫度評分中，多維度型投資溫度評分計算方式絕大多數情形下比單維度型投資溫度評分好。在 12 個指數實驗結果中有 11 個是多維度型計算方式較好，只有在大中華區域的 MSCI Gold Dragon Index 相關性檢定實驗中係以單維度型投資溫度評分較好，但與多維度型投資溫度評分相關係數僅相差 0.002。

表一、各投資溫度評分與參考指數漲跌相關性檢定結果表

市場區域	參考指數	類別型	單維度	多維度
臺灣	台股加權指數	0.687	0.730	<u>0.839</u>
	MSCI Taiwan Index	0.705	0.741	<u>0.830</u>
美國	S&P500	0.564	0.615	<u>0.810</u>
	Nasdaq	0.574	0.615	<u>0.751</u>
大中華區域	FTSE Greater Chinese Index	0.725	0.752	<u>0.754</u>
	MSCI Gold Dragon Index	0.725	<u>0.751</u>	0.749
亞洲太平洋區域	MSCI Asia Pacific Index	0.688	0.721	<u>0.800</u>
	FTSE Asia Pacific Index	0.682	0.716	<u>0.798</u>
新興市場	MSCI Emerging Markets Index	0.659	0.684	<u>0.717</u>
	FTSE Emerging Index	0.649	0.673	<u>0.713</u>
歐洲	FTSE Europe Index	0.443	0.501	<u>0.717</u>
	MSCI Europe Index	0.391	0.459	<u>0.666</u>

2、臺灣區域實驗結果分析

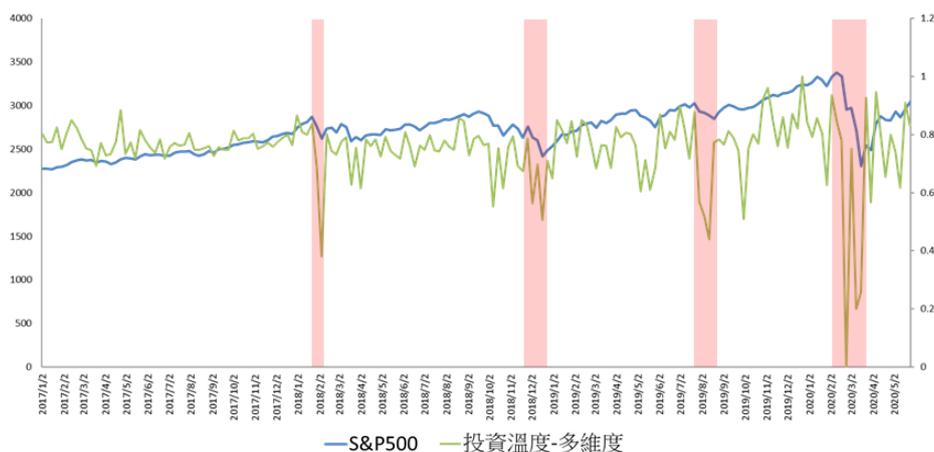
臺灣區域最佳相關係數為 0.839 為高度相關。相關性以折線圖方式表示如圖六，在下列時間點有很高的相關性：2018 年 2 月，股市漲幅過高、通膨預期升溫、公債殖利率上漲，引發程式交易避險，2 月 6 日台股創史上第 6 跌幅。2018 年 10 月，美國升息、預告 2019 年起對中國加徵關稅，10 月 11 日台股創史上最大跌點。2018 年 12 月，華為財務長遭捕，引發貿易戰，12 月跌幅 2.16%，為 2001 年來最差。2019 年 5 月，美中貿易戰升溫，5 月 14 日台股跌破年線。2019 年 8 月，川普下令美企撤出中國，8 月台股重挫逾 380 點。2020 年 2 月，新型冠狀肺炎造成全球經濟停滯。



圖三、台股加權指數與多維度型投資溫度相關圖

3、美國區域實驗結果分析

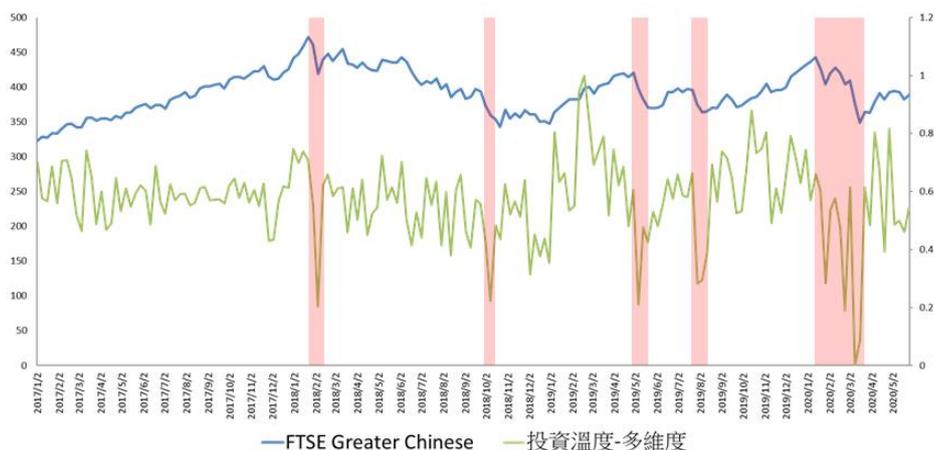
美國區域最佳相關係數為 0.810 為高度相關，相關性以折線圖方式表示如圖七，在下列時間點有很高的相關性：2018 年 2 月，2 月 4 日 5 日美股大跌 1,800 多點。2018 年 12 月，華為財務長孟晚舟遭捕，引發貿易戰戰火，12 月創 10 年來最糟單週表現。2019 年 8 月，川普下令美企撤出中國，美股在 8 月三度重挫至少 2.5% 以上。2020 年 2 月，新型冠狀肺炎造成全球經濟停滯。



圖四、S&P500 與多維度型投資溫度相關圖

4、大中華區域實驗結果分析

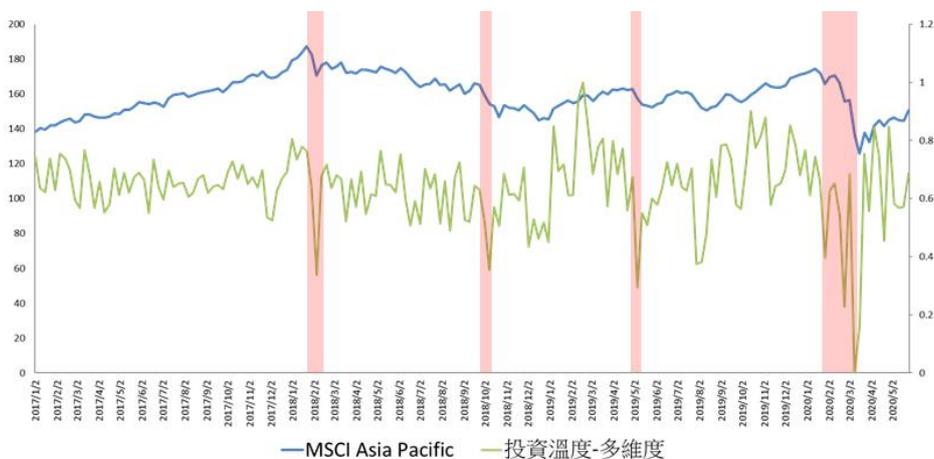
大中華區域最佳相關係數為 0.757 為高度相關。相關性以折線圖方式表示如圖八，在下列時間點有很高的相關性：2018 年 2 月，股市漲幅過高、通膨預期升溫、引發美股大跌，2 月 8 日港股開盤跌 800 點、上證 50 跌破半年線。2018 年 10 月，美國預告 2019 年起開始對中國加徵關稅，10 月上海 A 股跌幅 9.6%，港股跌幅 7.97%，台灣店頭市場指數跌幅 12.26%。2019 年 5 月，美中貿易戰對峙氛圍升溫，上證綜指 5 月創 2011 年以來最差的 5 月表現，港股 5 月下跌 2584 點或 8.7%。2019 年 8 月，川普下令美企撤出中國，港股 8 月中旬前下跌 2000 點，跌幅 7.6%。2020 年 1 月，新型冠狀肺炎造成中國大陸全面封城。



圖五、FTSE Greater China Index 與多維度型投資溫度相關圖

5、亞洲太平洋區域實驗結果分析

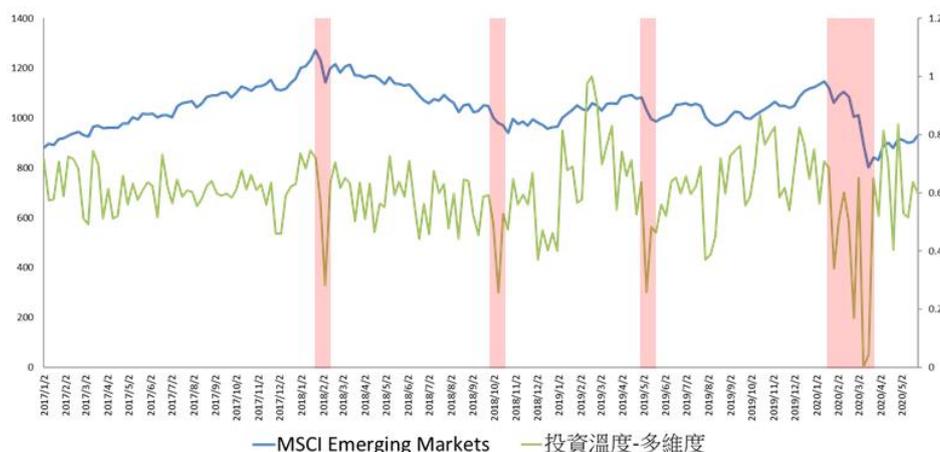
亞洲太平洋區域最佳相關係數為 0.800 為高度相關。相關性以折線圖方式表示如圖九，在下列時間點有很高的相關性：2018 年 2 月，股市漲幅過高、通膨預期升溫、引發美股大跌，2 月 6 日亞股趴成一片，上證指數收跌 4.05%，香港恆生指數跌逾 3%，日股跌逾 2%，韓股跌 1.8%，台股收跌 156 點。2018 年 10 月，美國預告 2019 年起開始對中國加徵關稅，10 月 11 日亞股哀鴻遍野，日經 225 指數收跌 3.89%；南韓股市收跌 4.44%，為 2011 年 11 月以來最大跌幅 2019 年 5 月，美國宣布對 2,000 億美元中國商品加徵懲罰性關稅，亞股 5 月 6 日重挫。2020 年 2 月，新型冠狀肺炎造成全球性經濟停滯。



圖六、MSCI Asia Pacific Index 與多維度型投資溫度相關圖

6、新興市場區域實驗結果分析

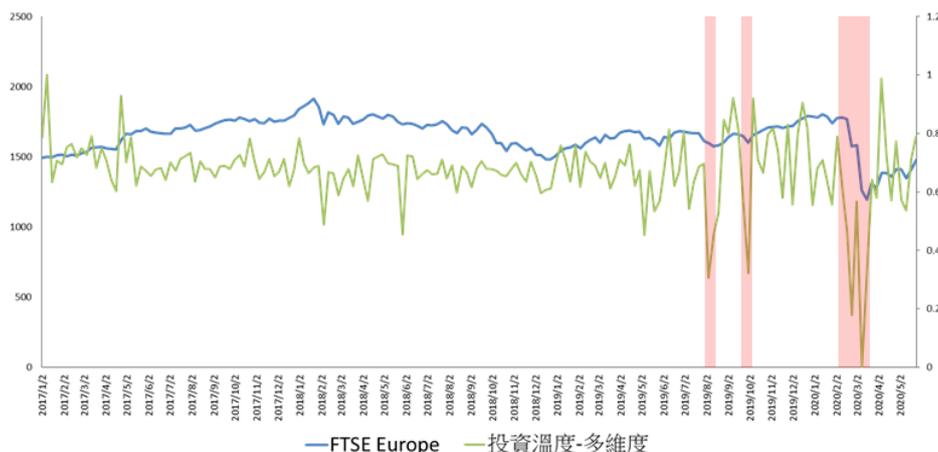
新興市場區域最佳相關係數為 0.724 為高度相關。相關性以折線圖方式表示如圖十，在下列時間點有很高的相關性：2018 年 2 月，股市漲幅過高、通膨預期升溫、引發美股大跌，恐慌氣氛導致新興市場資金顯著外流。2018 年 10 月，美國預告 2019 年起開始對中國加徵關稅，10 月新興市場股匯市表現慘烈。2019 年 5 月，美中貿易戰對峙氛圍升溫，新興市場 5 月中旬將 2019 年漲幅全部回吐。2020 年 2 月，新型冠狀肺炎造成全球性經濟停滯。



圖七、MSCI Emerging Markets Index 與多維度型投資溫度相關圖

7、歐洲區域實驗結果分析

歐洲區域最佳相關係數為 0.717 為高度相關。相關性以折線圖方式表示如圖十一，在下列時間點有很高的相關性：2019 年 8 月，川普下令美企撤出中國，美中貿易衝突再起，引發全球股市震盪，歐股也難逃修正壓力。2019 年 10 月，英國正式提出脫歐新方案，歐股全面大跌超過 3%。2020 年 2 月，新型冠狀肺炎造成全球性經濟停滯。



圖八、FTSE Europe Index 與多維度型投資溫度相關圖

五、結論及未來展望

本研究所做的實驗僅著重於驗證投資溫度評分計算方式與目標市場漲跌趨勢的相關性檢定，在投資溫度評分計算的方式上，本研究係採用類別型、單維度型及多維度型 3 種方式，實驗結果顯示，維度型投資溫度評分計算方式比類別型為好，而多維度型投資溫度評分計算方式在大多數的情形下比單維度型為好，在各組實驗結果相關係數平均值為 0.760，屬高度相關。未來展望及方向主要有幾個重點：

- (一)持續增加網路財經文本蒐集來源。本研究使用的文本來源為網路論壇及網路新聞，如能從其他財經網站或社群媒體中蒐集更多文本，對於將網路投資訊息量化為能夠代表投資大資心理的投資溫度評分更具客觀性。
- (二)本研究的目標是設定在國家及區域等大範圍市場，依據「行情=資金+心理」的投資法則，可以嘗試使用央行的貨幣供給總量、指數交易月均量、全球共同基金資金流向等資料做為「資金」面的變數，使用本研究的投資溫度評分做為「心理」面的變數，進而建構大範圍市場區域的金融商品薦購系統。
- (三)在網路財經文本情感分析的研究中，用來預測股價走勢是大多數研究者的目的，可以嘗試使用本研究投資溫度評分的計算方式，結合基本分析指標(總體經濟資料等)及技術分析指標(移動平均線、KD 值等)，用於建構市場趨勢預測模型。

參考文獻 [References]

- [1] 財團法人中華民國證券暨期貨市場發展基金會，*基金投資人投資行為與偏好問卷調查分析*(資產管理人才培育與產業發展基金委託專題研究)，P.31，民國 107 年。
- [2] 安德烈·科斯托蘭尼，*一個投機者的告白之金錢遊戲*(增修版)，民國 107 年。
- [3] B. Pang and L. Lee, *Opinion mining and sentiment analysis*. Foundations and Trends in Information Retrieval, 2008. 2(1–2): p. 1-135, 2008.
- [4] R. A. Calvo and Sidney. D'Mello, "Affect detection: An interdisciplinary review of models, methods, and their applications," *IEEE Trans. Affective Computing*, vol. 1, no. 1, pp. 18-37, 2010.
- [5] R. Feldman. "Techniques and applications for sentiment analysis," *Communications of the ACM*, vol. 56, no. 4, pp. 82-89, 2013.
- [6] Russell, J.A., *A circumplex model of affect*. Journal of personality and social psychology, 1980. 39(6): p. 1161, 1980.
- [7] Jin Wang, Liang-Chih Yu, K. Robert Lai and Xuejie Zhang, "Community-based Weighted Graph Model for Valence-Arousal Prediction of Affective Words," *IEEE/ACM Trans. Audio, Speech and Language Processing*, vol. 24, no. 11, pp. 1957-1968, 2016.
- [8] 蔡宇祥，*股市趨勢預測之研究－財經評論文本情感分析*，民國 104 年。
- [9] Qu et al.，*Quantifying Correlation between Financial News and Stocks*，IEEE Symposium Series on Computational Intelligence (SSCI)，2016.
- [10] 盧奕叡，*深度學習與情感分析應用於股價預測*，民國 107 年。
- [11] Zhao et al.，*Inferring private information from online news and searches: Correlation and prediction in Chinese stock market*，Physica A，2019.
- [12] Hwang et al.，*Interdependency between the Stock Market and Financial News*，IEEE International Conference on Big Data (Big Data)，2019.
- [13] Bourezk Hind et al.，*Analyzing Moroccan Stock Market using Machine Learning and Sentiment Analysis*，2020 1st International Conference on Innovative Research in

Applied Science, Engineering and Technology (IRASET) Innovative Research in Applied Science, Engineering and Technology (IRASET), 2020 1st International Conference on. :1-5 Apr, 2020.

- [14] M. M. Bradley and P. J. Lang, “Affective norms for English words (ANEW): Instruction manual and affective ratings,” Technical Report C-1, University of Florida, 1999.
- [15] A. B. Warriner, V. Kuperman, and M. Brysbaert, “Norms of valence, arousal, and dominance for 13,915 English lemmas,” *Behavior research methods*, vol. 45, no. 4, pp. 1191-1207, 2013.
- [16] M. M. Bradley and P. J. Lang, “Affective norms for English text (ANET): Affective ratings of text and instruction manual. Technical Report. D-1, University of Florida, 2007.
- [17] S. Buechel and U. Hahn. “EmoBank: Studying the impact of annotation perspective and representation format on dimensional emotion analysis,” in *Proc. of EACL*, pp. 578-585, 2017.
- [18] D. Tang, F. Wei, B. Qin, N. Yang, T. Liu, and M. Zhou, “Sentiment embeddings with applications to sentiment analysis,” *IEEE Trans. Knowledge and Data Engineering*, vol. 28, no. 2, pp. 496-509, 2016.
- [19] L. C. Yu, J. Wang, K. R. Lai, and X. J. Zhang, “Refining word embeddings using intensity scores for sentiment analysis,” *IEEE/ACM Trans. Audio, Speech and Language Processing*, vol. 26, no. 3, pp. 671-681, 2018.
- [20] N. Malandrakis, A. Potamianos, E. Iosif, and S. Narayanan, “Distributional semantic models for affective text analysis,” *IEEE Trans. Audio, Speech, and Language Processing*, vol. 21, no. 11, pp. 2379-2392, 2013.
- [21] G. Paltoglou and M. Thelwall, “Seeing stars of valence and arousal in blog posts,” *IEEE Trans. Affective Computing*, vol. 4, no. 1, pp. 116-123, 2013.
- [22] S. Du and X. Zhang, “Aicyber's system for IALP 2016 shared task: Character-enhanced word vectors and boosted neural networks,” in *Proc. of IALP*, pp. 161-163, 2016.
- [23] P. Goel, D. Kulshreshtha, P. Jain, and K. K. Shukla, “Prayas at EmoInt 2017: An ensemble of deep neural architectures for emotion intensity prediction in tweets,” in *Proc.*

of *WASSA*, pp. 58-65, 2017.

- [24] J. Wang, L. C. Yu, K. R. Lai and X. Zhang, "Tree-Structured Regional CNN-LSTM Model for Dimensional Sentiment Analysis," *IEEE/ACM Trans. Audio, Speech and Language Processing*, vol. 28, pp. 581-591, 2020.
- [25] L. C. Yu, J. Wang, K. R. Lai and X. Zhang, "Pipelined Neural Networks for Phrase-level Sentiment Intensity Prediction," *IEEE Trans. Affective Computing*, to appear.
- [26] L. C. Yu, L. H. Lee, S. Hao, J. Wang, Y. He, J. Hu, K. R. Lai and X. Zhang, "Building Chinese Affective Resources in Valence-Arousal Dimensions," in *Proc. of NAACL-HLT-16*, pp. 540-545, 2016.