

日文歌曲情緒分類

作者：Ian Hsieh

➤ 專題簡介

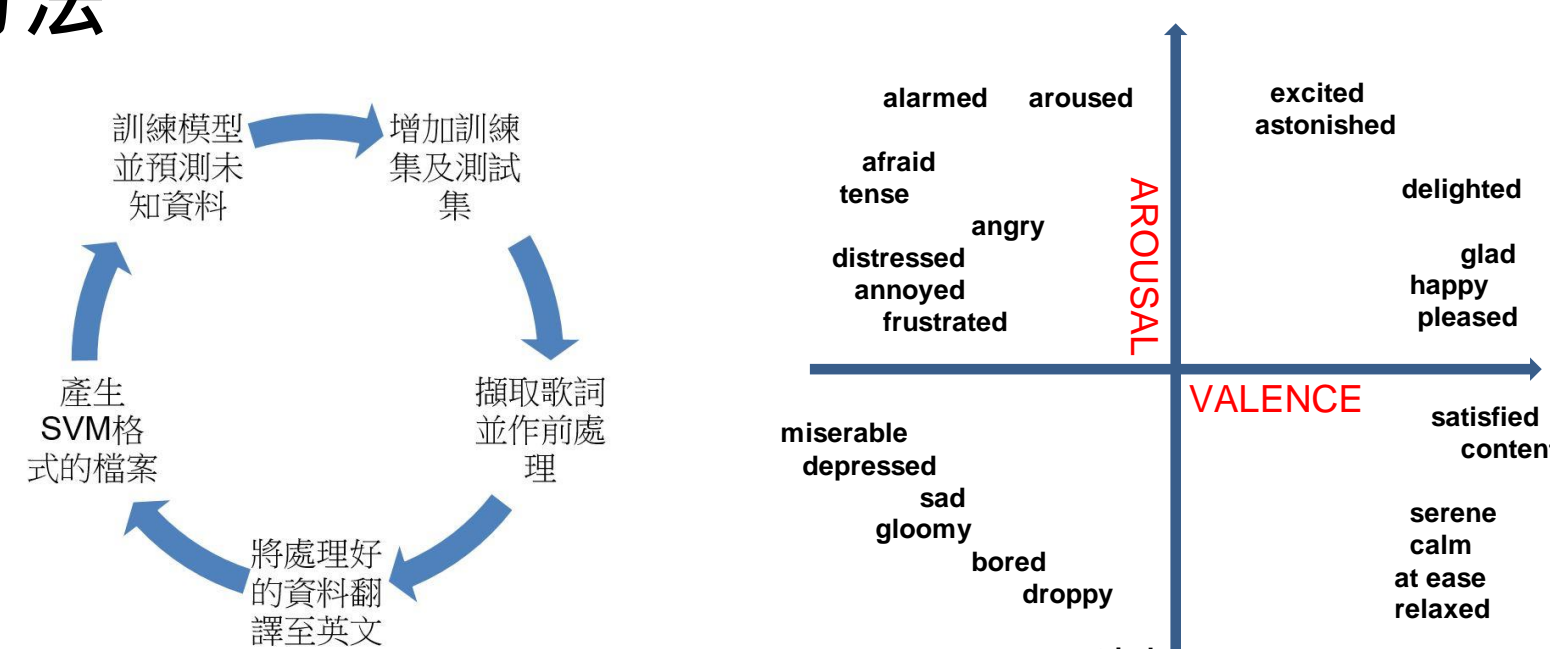
此專題嘗試從機器翻譯的面向來探討：「日文歌曲是否也能取得有力的情緒分類系統。」關於歌曲情緒分類之研究，從音訊分析的角度切入，近年來已大致飽和。歌曲除了旋律之外，歌詞也被證明為一項分類的重大指標。在中英文上被廣泛研究，且測試過跨國文化對歌詞不同觀感的情況下[1]，此專題嘗試從中文語系實驗者的身分，發掘日文與英文歌詞間對於歌曲情緒是否能達成一致的看法。

本專題的系統先取得日文歌曲歌詞，資料數量共125筆，100筆作為訓練集，25筆作為測試集。我們提供了歌手以及歌名來做為自動取得歌詞的依據，再依靠人工的方式，根據Russell's model[2]，以valence-arousal為兩軸的四大象限來做分類。接下來進行文字前處理的動作，因著日文歌詞中含有的外來語大多為英文，故選擇使用翻譯的方式來判斷情緒分類，而翻譯為英文單詞時也會較為準確，最後以自動程式來完成多筆資料翻譯。

而機器學習便是此專題之主要方法，利用Support Vector Machine的特性，將歌詞切分為多個feature，計算每項feature的權重，並以自定義的四個分類作為歌詞的Gold standard，也就是我們想找出的分類，將這份資料訓練出一個模型，再將沒有Gold standard的資料做為測試集，利用先前訓練出的模型套入這份資料之中，以辨別這個模型在歌曲情緒分類上的表現。而在實驗過後也能看出兩個領域趨勢，一為跨語系對於情緒分類之影響，以及歌詞與情緒分類之間的關聯性，最後可將此二趨勢融合分析，尋找未來的研究改進方向。

➤ 專題成果

實驗方法



流程參照左上之圖示，首先是從歌詞網站(mojim.com)取得歌詞，使用Html Agility Pack在C#中實作程式來自動化此過程，網站之XPath時常更換，故須定時檢驗才能得到可用之資料。在取得資料後，便以人工的方式進行四大類型的標註，右上之圖示為Russell之歌曲分類模型，分別Arousal部分時以X軸為界，而分別Valence的部分時以Y軸為界。第一象限便是激昂、正向情緒的類別，第二象限則是激昂、負面情緒的類別，第三象限是冷靜、負面情緒的類別，而第四象限便是冷靜、正向情緒的類別，主要以歌詞為主、音訊為輔的判別方式。

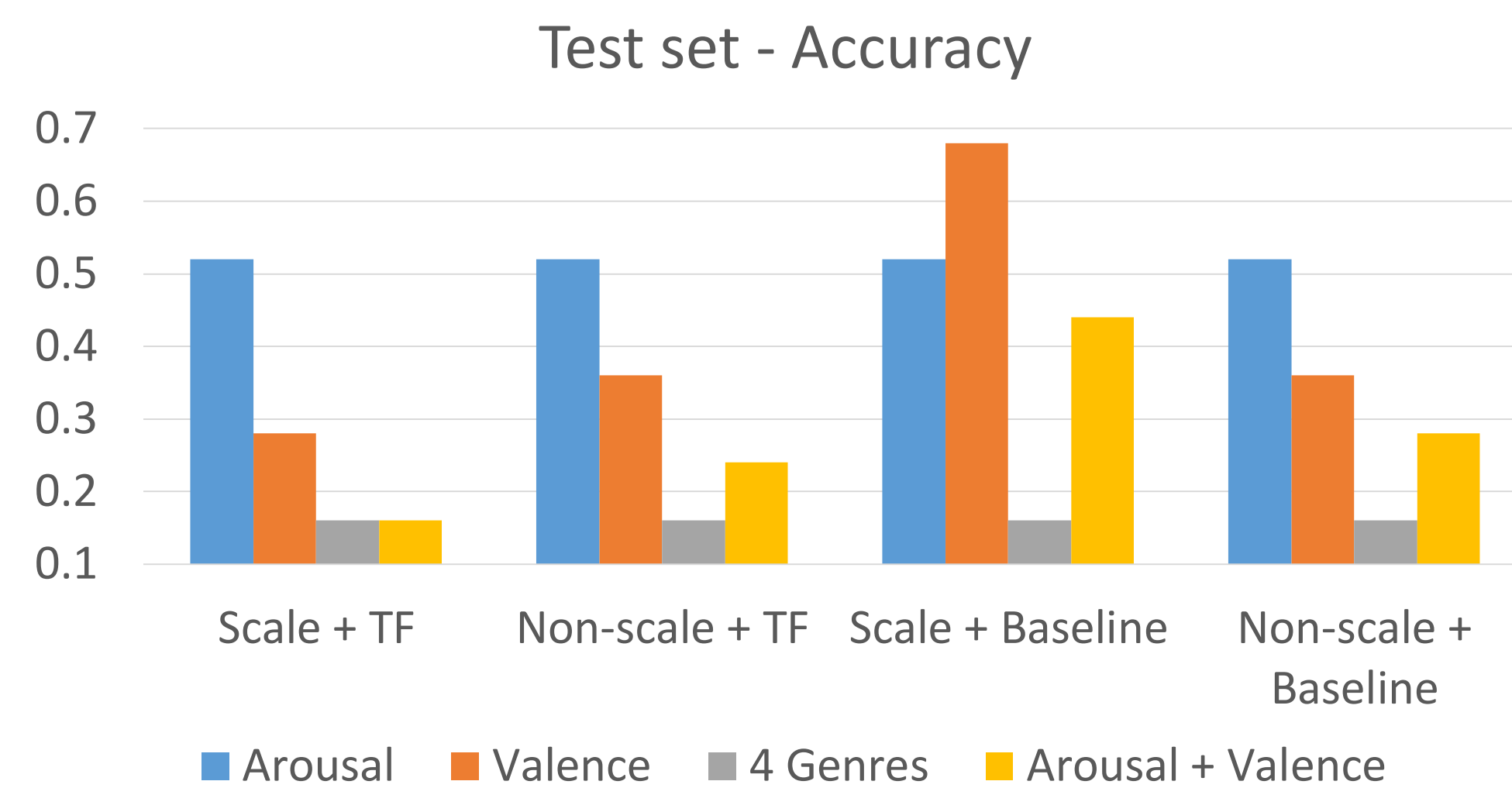
希望 + Hope
の Of
道 The way
は To
は End
は The
は Do not have
ない Neg

此專題在SVM訓練時主要採用word-level feature，並經實驗後發現全文bag-of-word(unigram)比全文bigram來得更加有預測性，此事實正可驗證現今歌詞的多樣性。故在前處理時，除了第一步的日文斷詞(kuromoji)及網頁擷取雜訊去除等，在翻譯斷詞資訊後，必須對該英文單詞作出Stemming及Lemmatization的正規化步驟，將字詞還原至原本的字根，如is, are等詞可還原至be，Going, went等詞可還原至go，並將Stopwords如the, at, of等介係詞及難以幫助分類的字去除。爾後再對否定詞進行bigram的包裝，將否定詞前後的一個單詞與否定詞本身合併成一個feature，上方左圖便為LIBSVM之檔案格式，行首為分類標註，而後面依序為「[Feature Number]: [Feature Weight]」，以此類推。

$$WFO(t, c_i) = P(t|c_i) \lambda \left\{ \max \left(0, \log \frac{P(t|c_i)}{P(t|\bar{c}_i)} \right) \right\}^{1-\lambda}$$

而在對於Feature權重計算的部分，採用了兩種方法，一為Baseline方法，尋找正向及負向之情緒字典[4][5]針對有情緒的字加重權重，而中性字眼則採歌詞情緒字眼權重總合之平均，可參照上一段之說明圖示；二為基於Term-frequency對所有字進行之算法處理，詳細公式可參考上圖之Weighted Frequency and Odds[3]， $\lambda=0$ 時會有最佳表現，原用於評論分類之文字處理。

實驗成果



以上共有四種不同方法針對四個不同的狀況作出預測模型，上圖為建立在十二種模型下進行未知測試資料的預測。Scale為將feature之權重範圍縮小在[-1, 1]之間，TF及Baseline則是上述提過的兩種權重測量方法；Arousal及Valence是將四大分類分作兩項兩大分類處理，4 Genres則是直接處理四大分類，Arousal + Valence則是綜合兩個模型的結果，故沒有以此方法獨立訓練模型。

從結果可以發現一些趨勢，此專題的方法更適用於Scale過的資料；Scale + Baseline與Arousal + Valence所作的四大分類預測會有最好的表現，遠比4 Genres直接分類的方法好上許多；Arousal及Valence表現的部分，則可看出歌詞較能預測Valence的部分，而對於Arousal則關聯較小；在Test部分普遍低落的情形，則顯現了整體訓練資料不足的問題，在第一及第四象限的歌詞遠比第二及第三象限來得多，而測試資料卻大多是第二及第三象限的歌詞；在情緒字典及翻譯的部分也可以看出需要擴充及提高正確性來幫助日新月異且多樣性高的歌詞，否則日文歌詞一字多義的情形會使情緒更難以預測；權重算法依然也是可以最佳化的一部份，來幫助SVM發揮最大的效能。

參考論文

- [1] Abhishek Singhi and Daniel G. Brown, On Cultural, Textual And Experiential Aspects Of Music Mood, ISMIR 2014.
- [2] Xiao Hu and J. Stephen Downie, When Lyrics Outperform Audio For Music Mood Classification: A Feature Analysis, ISMIR 2010.
- [3] Shoushan Li, Sophia Yat Mei Lee, Ying Chen, Chu-Ren Huang and Guodong Zhou, Sentiment Classification And Polarity Shifting, Coling 2010.
- [4] Mingqing Hu and Bing Liu, Mining And Summarizing Customer Reviews, KDD 2004.
- [5] Bing Liu, Mingqing Hu and Junsheng Cheng, Opinion Observer: Analyzing And Comparing Opinions On The Web, WWW 2005.