

1. INTRODUCTION

1.1. Background

According to Oxford's Advance Learner's Dictionary, news is a reporting of new information about something that has happened recently. Generally, it can be an event or fact that is reported formally. In the early days, news is being delivered through traditional approaches such as radio channels and newspapers. The rapid development of internet and web has brought about a change in the form of news publication. Online news platforms are springing up like mushroom after the rain. With the mobile devices made popularised, online news has become more and more easily accessible.

The massive publication of news is believed to have trigger some feelings or sentiment in the readers be they paper-based or digital. People tend to have their feeling expressed by talking and debating the topic of interest at the coffee shop on in the neighbourhood park during the days when the internet has not been widely used yet. Such scenario of offline discussion is often emotionally oriented and still can be seen anywhere until today. In this digital age, the online platform also provides another alternative for the expression of thoughts and feelings. Some of these online platforms are social media sites and blogs such as Twitter, Facebook and Blogger where the online user usually express their feelings in text or comment.

Today, one of the interesting ways for gathering feedback from online users is the emotion voting system. This special feature understands the reader's view or attitude towards a situation or a site content with emotion representation. This work presents topic related to sentiment analysis and to find out an approach for predicting readers' sentiment on the news by utilising the emotion voting feature available based on multi-label classification.

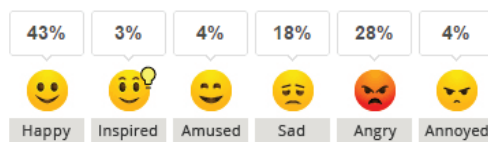


Figure 1.1 Emotion Voting System from TheStar online¹

¹ <http://www.thestar.com.my/News/Nation/2015/10/30/Vaping-Health-Ministry-effects/>

In fact, sentiment analysis is one of the three key type of big data analytics, which is the predictive analytics as the sentiment polarity (positive or negative) is the set of data that researchers do not have when doing the prediction (Bertolucci, 2013). This means that raw data are usually processed to generate more valuable details by using the right set of tools and techniques. Meanwhile, the study of emotion classification is focused on two perspectives, which are the writers' perspective and the readers' perspective.

A different approach is used to perform the prediction and classification depending on the perspective that one is looking at. Some of these approaches adopted in past researches are the knowledge-based approach, language models as well as discourse structures and semantics. Both supervised and unsupervised methods are being implemented in sentiment prediction for classification. Supervised methods suggested high demands of labelled training data and usually recorded better performance compared to an unsupervised method whereby the acquisition of unlabelled data is fairly effortless.

1.2. Motivation

Emotions are often one of the influencing factors during decision making, particularly in the traditional society as people tend to seek for opinions as a reference. Knowing what others think and feel is essential for various group of people in the society, regardless of which occupational area they are engaged in. They can be marketers, politicians, public relation officials, managers or even general people. Marketer and advertiser can track public response to an advertising campaign; politician and public relation officials are able to get the insight of public emotions to the certain event especially on the political situation of a country; stock traders are able to capture the market behaviours for stock trading and trend analyser can follow the latest technology and entertainment trends.

Also, news domains contents are huge in general and are often categorised into many different aspects such as world news, nation news, entertainment news, sports news, financial news, technology news and much more. Nevertheless, it is quite impractical to consume all of its content with such a huge domain in order to perform a Media Response Analysis. Seeing the needs to improve the existing news categorisation, it is worth questioning if there are other ways for enhancing the news browsing experience with a different perspective.

The study of emotion prediction aims to explore the sentiments triggered by affective text which seems to be a lot harder to be done on news contents as they often do not come with readers' opinions or reviews which tend to be more subjective. In recent years, numerous news websites have introduced the emotion voting features allowing the expression of feeling towards any topic hosted by the news portal. Therefore, the availability of such data facilitates the research on sentiment prediction and hence encourage the emotion classification of news article.

1.3. Research Problem

With the overwhelming growth of online news portal or services, sentiment analysis has become increasingly important, especially in the field of media monitoring (Scholz & Conrad, 2012). The conventional methods of gathering readers' opinions through interviews or questionnaires are time consuming and require a lot of human effort as news articles may span larger subject domains (Magadza, et al., 2014; Balahur, et al., 2010).

Sentiment analysis or opinion mining in blog articles or product review can be very different compared to news article as the content are typically subjected to a particular topic or issue concerned (Scholz & Conrad, 2013). Meanwhile, news articles sentiment analysis turned out to be more challenging due to the need of journalist to remain neutral by avoiding from writing content in a tone that may sparks unpleasant feelings.

Current research works focuses more on sentiment analysis by extracting opinions on specific domains that contain high volume of subjective text, such as product or movie reviews. Comparatively, sentiment prediction is more difficult to carry out on news text for the reason that news editor usually presents the event in an objective manner in the news reports. Thus, opinion and emotions are not significant as they are being transmitted implicitly.

Moreover, the ways of sentiment expression can differ correspondingly depending on the sentence position in news content. For example, news headlines should have higher sentiment intensity compared to news bodies in order to catch readers' attention. Most state-of-art work on sentiment analysis also suggested the approach of adjectives as an

indicator of sentiments (Hatzivassiloglou & McKeown, 1997; Hatzivassiloglou & Wiebe, 2000).

In real world setting, it is believed that news can trigger certain emotion in every mankind, be they trivia or breaking news. Typically, journalists will avoid using vocabulary that are clearly negative or positive, but attempting to express opinions by omitting some facts and highlighting others (Balahur, et al., 2010). Therefore, the same incident reported from different media platforms can possess various versions of reporting styles which in turn can lead to different degree of emotional charge as humans are emotionally driven and reader's emotions aroused by the text content do not necessarily align with those of writers. All of these accounted for the further research on the proposed title and method as a result.

1.4. Objectives of the Research

1. To determine an appropriate word classes (nouns, verbs, adjective) in learning the model for classifying emotion label for news.
2. To determine the type of news contents (headlines, first paragraph, last paragraph) that are crucial in classifying the news emotion.
3. To label emotions for news from the same category (Nation, World, Sport or Regional).

1.5. Assumptions and Constraints

Only the readers' perspective will be taken into consideration in this research, intending to understand and explore the thoughts of readers on the news. In other words, the intention of writers is not going to be the primary focus in this study. It is assumed that there is no repetitive vote recorded on the emotion label of a particular news article. Besides, the data collected will not only focus on a single emotion with the most significant numbers of votes as humans can have mixed feeling towards an event. Therefore, multiple emotions will be captured to show a more accurate feeling reflected by the readers on the particular topic he or she has read.

It is also assumed that the model proposed for news classification based on emotion labels is only applicable for data in Standard English. This language constraint suggests that the application of news portal in other languages such as Chinese, Korean or Japanese

may give inaccurate results of sentiment prediction and affect the performance of emotion classification. Another constraint is that the platform used for collecting the readers' sentiment is targeted on Malaysian local news portal whereby the emotion voting feature is still yet to be a common trend in the meantime.

1.6. Scope of Research

In this research, The Star Online is used as data source for news article. The emotion label feature found in it plays a pivotal role as our research data whereby each emotion depicted the feelings or impression of the readers to the news content. This collection of sentiment data is somewhat valuable as they provide pieces of information that tell the public general feeling towards a specific topic or content. Therefore, the emotion classification discussed in this research is putting primary focus on the reader's perspective.

Supervised method of classification is implemented to train emotion labelled data in this research. To obtain a more comprehensive set of data for learning and analysis purpose, the data mining is not just on news headline but also considering the body of the news. By using the part-of-speech tagger, the adjective and adverbs can be identified to analyse their emotional tendency.

With the specific representation of the emotions, a more relevant model is proposed in the hope of expanding the contribution of sentiment classification for news so as to introduce a more interesting approach of reading news by classifying news based on sentiments instead of the typical categories.

1.7. Outline of Thesis

The rest of this paper is organised as follows:

- (i) Chapter 2 discusses the related work for sentiment analysis and classification as well as the overview of different techniques and approaches implemented in these field.
- (ii) Chapter 3 describes the proposed methodology for classifying news sentiment based on multiple emotions captured through the emotion voting system.

- (iii) Chapter 4 presents a brief data implementation and the planned evaluation of the corpus for the research.
- (iv) Chapter 5 presents the explained the performance metrics use to measure the outcome of the research as well as analysis and discussion on the findings.
- (v) Chapter 6 concludes the final thesis with brief summary on its finding, research contribution and future works.