Colour Coded Emotion Classification in Mental Health Social Media

Neil Vaughan
University of Chester,
Thornton Science Park,
Pool Ln, Chester, CH2 4NU
n.vaughan@chester.ac.uk

Maurice Mulvenna
Ulster University
Shore Road, Newtownabbey, Co.
Antrim, BT37 0QB
md.mulvenna@ulster.ac.uk

Raymond Bond
Ulster University
Shore Road, Newtownabbey, Co.
Antrim, BT37 0QB
rb.bond@ulster.ac.uk

This research applies emotion detection to messages from online mental health social media. In particular, this focusses on specialised social media for users to report health or mental health problems. Automatically detecting the emotion in social media can help to rapidly identify any concerning problems which could benefit from intervention aiming to prevent self-harming or suicide. Detecting emotions enables messages to be colour coordinated according to the emotion to enhance the human-computer interaction. A supervised classification method is applied to a labelled dataset and results presented. A prototype user interface system is developed based on detecting emotion, colour coding the message to display detected emotions to users in real-time.

Emotion detection. Social Media Analysis. Text mining.

1. INTRODUCTION

Mental health problems are important affecting one in four citizens at some point during their lives and too often lead to suicide (EU Green Paper 2005). Social networks continually provide several functions for individuals and contribute towards networking within the social structure. Online social media interaction creates large online datasets of messages such as Facebook or tweets. More specialised social networks also provide for more specific functions and interests. Social networks also exist to support people experiencing problems or difficulties including health and mental health. One such social network is talklife which provides an app giving access to a safe social network for people who want to get help and advice. This support network is targeted towards helping people who are experiencing problems in four categories: Self-harming (including suicide), Sexual, Violence or Dieting.

These social networks occasionally receive messages containing a range of emotions from the emotional spectrum. There is an intention by the monitors of the social network to detect any extremely negative emotional messages, and respond to those rapidly. These negative emotions may indicate or lead towards a higher risk of self-harming and suicide and could be preventable if a counteraction is put into place early.

The elements of human-computer interaction are considered as the messages are coloured. Colouring messages according to the emotions could enable users to easily see the emotions of each message by the colour of the background.

In order to detect emotion, this paper presents experimental methods for emotion detection in text. Emotion mining is the field of natural language processing which is concerned with the detection and classification. It is a subfield of semantic analysis which contains emotion and opinion mining.

Both tasks depend on an emotion model to classify detected emotions and to associate a colour depending on the location of detected emotion.

This research paper demonstrates preliminary results from two tasks: (1) classification of the emotion within anonymised social media messages from talklife to assign emotion labels, and (2) adding colour coordination so that the emotion is easily visible to social media users or moderators for usability purposes, which is demonstrated through a designed prototype visualising how the colours can be implemented beneficially in a human-computer interface.

2. BACKGROUND OF TEXT PROCESSING

The first milestones in natural language processing (NLP) began around 1950 when rapid progress was being made in machine learning (Lehnert & Ringle, 1982). The first NLP techniques extracted only single words and interpreted them separately. These were useful for language translation.

Published by BCS Learning & Development Ltd.
Proceedings of the British HCI 2018, Belfast, UK

Limitations of methods using individual words included the inability to correctly understand words which can have opposite meaning depending on the surrounding words and structure. An example is the word “accident” which can be good or bad (Cambria & White, 2014). Also the phrase “not happy” could be detected as positive if the algorithm only focusses on the word happy.

Recent approaches to NLP for sentiment analysis involve training large neural networks with large knowledge bases of vocabulary. One such method in this approach is called ‘skip-gram’ which passes each keyword to another neural network, which then predicts the words either side to produce a binary tree which can be analysed (Witten et al., 2016).

Recent SemEval winning methods have shown that word embedding is shown to perform best for sentiment analysis. The topic was well explored in the Computational Linguistics community, with machine learning (Strapparava & Mihalcea 2008), using a Lexicon to associate colour (Volkova et al., 2012), crowdsourcing Word-Emotion associations (Mohammad & Turney, 2013), Word–Colour associations (Mohammad, 2011), and color of text emotions (Strapparava & Ozbal 2010).

The most machine learning and ‘deep learning’ languages for NLP are Python, R, and Java (Puget, 2016). Machine learning has been used with various NLP techniques such as ‘bag-of-words’ which analyses each word separately without context can be used with machine learning (Cronin et al., 2017). Machine learning was used to perform emotion mining by Alm et al., (2005) using emotion from text to change how words were spoken by a text-to-speech system.

NLP is only one aspect of the project, with the display of mood colours being the other. Mapping certain colours to moods will be different for each user so groups of colours will have to be assigned to each mood using the most common colours as a starting map (Moon, Kim, Lee, & Kim, 2013).

The first NLP systems attempted to parse text using ‘semantic information processing’ which uses keywords in sentences to trigger actions (Lehnert & Ringle, 1982). Early programs used a statistical model called n-grams where n is the number of probabilities required to specify a statistical model (Shannon, 1948). One such program was called SHRDLU, which was able to determine actions specified by a user using natural language (Winograd, 1972).

From this syntax based approach, research into NLP split into using semantic analysis and machine learning techniques (Cambria & White, 2014). Semantic analysis aims to recognise the semantic structure of a sentence to understand its meaning. “… in order to understand a sentence, it is necessary to know its syntactic pattern.” (Chomsky, 1957). This approach was popular in machine translation as the second language could be generated over the semantic structure of the original sentence (Schank, 2014).

Current approaches to NLP utilise neural networks for machine learning on top of adaptable knowledge bases. Current research into machine translation focuses on recurrent neural networks with enhanced ‘long short-term memory’ to better maintain information throughout a sentence (Hirschberg & Manning, 2015). This memory is achieved via backpropagation through time (BTT) algorithms which allow data to be propagated back through time and be remembered over multiple steps in the hidden layer of the network (Mikolov et al., 2011).

2.1 Emotion Mining

Current emotion mining techniques work on the sentence level, classifying emotion for each sentence, and use annotated data models to calculate the expressed emotion (Yadollahi et al., 2017). Polarity determination is also important for emotion mining as it determines whether a sentence is expressing positive or negative emotions (Ravi & Ravi, 2015). This is useful when applied to naïve approaches which look for key emotion words instead of analysing the entire sentence.

For this project, the aim was to classify the emotion that the writer aims to invoke in the user. Pizzi. et al, (2007) expands this and used NLP to present each character in a story with their own internal emotion to better capture the emotions between characters. This can be used to calculate the overall mood of the section with weights given main characters.

Alm, et al., (2005) also used narrative text as the basis for their research. Their application of emotion mining is used to enhance a text-to-speech system for reading fairy tales. Machine learning was also utilised to detect the valence of basic emotions to change the pitch and speed of the output speech.

2.2 Mapping Colour to Emotion

To detect emotions, they must first be categorised in a model so that it can be analysed against the text. An emotional model is a system which uses either category of emotions, such as ‘anger’, or emotional dimensions, such as valence and arousal (Burkhardt & Stegmann, 2009).

In a study to display the mood of music via coloured lighting, Moon, et al., (2013) used Thayer's emotion model to determine the emotion. The Thayer emotional model is a 2D grid of emotions plotted against arousal and tension (Thayer, 1990). Because this model doesn’t just use emotion adjectives, there is less ambiguity (Moon, Kim, Lee, & Kim, 2013).

The emotion a person associated with a colour can depend on many factors (Manav, 2017) including personal experience, memories, and cultural perceptions. A study into emotion and colour preferences from Ou, et al., (2004) found that some emotions may be associated with the same colour across many countries.

2.3 Suicide note text analysis

The first research into computational analysis of self-harming or suicide notes was by Pestian et al. (2010). Since then, automatic systems can outperform mental health professionals in separating genuine from fake notes. Suicide note analysis is a useful tool towards suicide prevention (Desmet and Hoste, 2013).

Supervised machine learning approaches rely on labelled data, of which various techniques have been applied: support vector machines, naive Bayes, Hidden Markov Models and memory-based learning (Banea et al., 2008), (Pak & Paroubek, 2010), (Wilson, Wiebe, & Hoffmann, 2005), (Rentoumi et al., 2010).

3. METHODS

The text mining method was applied to anonymised messages from the talklife social media app.

3.1 Dataset from talklife

The talklife dataset was generated by recording all talklife messages over a 12 hour period. This contains over 4600 messages; a new message is posted approximately every 9 seconds. The total messages in the dataset was reduced to 2665 once unsuitable messages were removed as described below. The quantity of data available is important because there is a clear correlation between performance and data availability (Desmet and Hoste, 2013). In recent years with increased big data, automatic techniques are crucial for analyzing large amounts of online text.

All messages were anonymised by removing the name of the user who posted the message and any likes or replies. The ethical implications of using sensitive talklife messages was reduced since the messages were already posted in a public place, and are available for anyone in the public domain to read online or through the free talklife app.

Text analysis on this type on online blog has various complexities: messages tend to contain a lot of slang words, phrases are used out of their usual context. Messages containing text over multiple lines were processed to remove the carriage returns. Some messages were longer than the 300 character limit and were truncated to the last complete word but missed the end of the sentence. Some messages didn't contain any words - these were removed from the test dataset. Many messages contain non-text ASCII characters which remained in the dataset if ASCII-255 compatible. Several messages contain text copied from musical lyrics. There were some foreign or illegible messages which were removed. Several messages were observed with spelling errors or unusual acronyms. When applying NLP modules on data with many spelling mistakes, they fail to cope with the intense surface variation and fail during early stages of analysis (Desmet and Hoste, 2013). Pre-processing steps to fix typographical errors can improve text analysis (Liu, 2010).

The talklife dataset is already effectively self-labelled which was an important aspect. Each user when posting a message has to select from a list of 41 emotion labels. The emotion list includes: Heartbroken, Sad, lonely, Depressed, Stressed, Confused and others as shown in Table 1. A problem with self-labelled data is that the emotion label selected by the user may not always accurately represent the emotion that appears to be within the message text itself, for example some users may select a label from the list at random when posting a message. It was found that the most frequent emotions posted during this timeframe were dominated by negative emotions: Sad (375), Meh (262), Lonely (221), Heartbroken (189), Tired (183), Calm (111), Anxious (105). The label 'Meh' expresses a lack of interest or enthusiasm, interpreted to mean a miscellaneous label for messages which the talklife user couldn't fit into another category.

***Table 1**. The frequency of forty-one emotion classes that talklife users self-labelled their messages with.*

No.	Label	Count	No.	Label	Count
1	Sick	17	20	Stressed	40
2	Angry	22	21	Confused	66
3	Inspired	28	22	Nervous	44
4	Surprised	5	23	Exhausted	87
5	Irritated	20	24	Numb	81
6	Furious	10	25	Meh	262
7	Encouraged	18	26	Happy	41
8	Caring	22	27	Frustrated	57
9	Annoyed	48	28	Worried	33
10	Loving	44	29	Hungry	29
11	Amazed	12	30	Insecure	72
12	Tired	183	31	Chilled	46
13	Lonely	221	32	Positive	61
14	Afraid	64	33	Supportive	84
15	Excited	11	34	Proud	18
16	Amused	67	35	Relaxed	32
17	Embarrassed	15	36	Calm	111
18	Heartbroken	189	37	Anxious	105
19	Sad	375	38	Playful	25

3.2 Classification methods

The classification method applied was based on previous research by Kiritchenko et al. (2014) and the AffectiveTweets package for analyzing emotion

and sentiment. These methods built upon our previous implementation for analysing messages on Twitter (Harvey et al., 2018).

First a pre-processing filter was applied, converting the text string to Sparse Feature Vectors (SFV). The SFVs are calculated including word and character n-grams. This has been previously useful for filtering out infrequent features and setting the weighting approach. A support vector machine (SVM) was trained. For comparison, SVM training was completed twice for each dataset, once with SFV pre-processed data and once with raw data. Ten-fold cross validation was applied to assess the classification accuracy which has advantages over using a training/test data split.

A variety of additional classification method were applied including applied 0-R classifier, Stacking, CVParameterStacking (Kohavi, 1995) and Support Vector Machine (SVM) which previously gave good results in our previous test (Harvey et al., 2018).

3.3 Emotion Detection Results

The results showed a production of 14% classification accuracy. This was achieved by classifying all instances as "sad" due to the dataset being skewed containing larger numbers in certain classes. Lower classification accuracy for the talklife dataset in comparison to previous results indicates higher complexity, potentially due to a larger search space due to 38 separate emotion class labels. Also this task is more complex due to the talklife data having less reliable labelling as each label was assigned by the user who submitted the message and each user may use labels based on their own interpretations.

In future work this talklife dataset could be re-labelled by an annotator to ensure the class labels are consistent and representative of the message wording. In previous research (Harvey et al., 2018) we applied the methods to a tweets dataset pre-processed into Sparse Feature Vectors (SFV), the trained SVM classified 74% of tweets correctly into one of the three labelled classes. Without any pre-processing, SVM classification was not as successful, producing 36% correct classification.

Future work could investigate various other classification methods successfully applied to emotion detection, including affective lexicons, training deep learning models or training a convolution Neural Network. Also the dataset could be adjusted into a binary label of emotional or neutral, which would simplify the classification search space.

3.3 Prototype App Colour Labels

In order to visually display the detected emotional content of messages to users or moderators, the next objective is to develop a prototype for the visual indication of emotion in messages. In the prototype, colours could be associated with the 38 emotional classes. Designs were created for how this could be integrated into a mental health social media within the web browser or mobile app. In future work we plan to implement this feature for use in a web browser to display emotion for real-time social media messages. The goal would be to use an individual web-page or use a web browser extension to add elements and styling. Figure 1 shows one of the proposed designs for integrating with colours in a mobile app for social media.

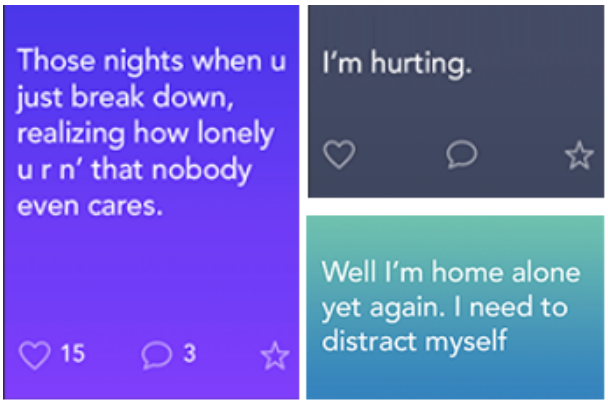

Figure 1: *Colour coordinated emotional messages from social media platform for mental health (talklife, 2018).*

4. CONCLUSION

The text mining method has been applied to talklife social media data for mental health. The detection results were previously shown to provide good classification results (Harvey et al., 2018). This work has extended to apply the technique for mental health related social media. A prototype web visualisation is proposed to improve human computer interaction and easily visualise emotion from colours.

This paper builds on the existing work in emotion modelling. A proof-of-concept emotion classifier was applied to the new dataset and a prototype web interface could be developed to display the resulting colours based on emotions. This could help to highlight any messages containing extreme emotions or indications of self-harming behaviours.

Future work could include extending this to other platforms. In future work, the user may be able to select which colours are associated with which emotions. Individuals have their own opinions about which colours relate to which emotions and it depends on each user's own personal background experiences.

5. REFERENCES

Alm, C. O., Roth, D., & Sproat, R. (2005). Emotions from text: machine learning for text-based emotion prediction. Proceedings of the conference on Human Language Technology and Empirical Methods in Natural Language Processing - HLT '05, 579-586.

Bantum, E. O., Elhadad, N., Owen, J. E., Zhang, S., Golant, M., Buzaglo, J., Giese-Davis, J. (2017). Machine Learning for Identifying Emotional Expression in Text: Improving the Accuracy of Established Methods. Journal of Technology in Behavioral Science, 2(1), 21-27.

Burkhardt, F., & Stegmann, J. (2009). Emotional speech synthesis: Applications, history, and possible future. Proc. ESSV.

Cambria, E., & White, B. (2014, May). Jumping NLP Curves: A Review of Natural Language Processing Research. IEEE Computational Intelligence Magazine, 9(2), 48-57.

Chomsky, N. (1957). Logical Structures In Language. Journal of the Association for Information Science, 8(4), 284-2910.

Desmet, B., & Hoste, V. (2013). Emotion detection in suicide notes. Expert Systems with Applications, 40(16), 6351-6358.

EU Green Paper. (2005). Improving the Mental Health of the Population. Towards a Strategy on Mental Health for the European Union. World Health Organization Regional Office for Europe.

Harvey R, Muncey A, Vaughan N, (2018) Associating Colours with Emotions Detected in Social Media Tweets, Symposium for Emotion Modelling and Detection in Social Media and Online Interaction, Proceedings of Artificial Intelligence and Simulation of Behaviour.

Hirschberg, J., & Manning, C. D. (2015). Advances in Natural Language Processing. Science, 349(6245), 261-266.

Kohavi, R. (1995). Wrappers for performance enhancement and oblivious decision graphs. Carnegie-mellon University, Pittsburgh PA.

Lehnert, W. G., & Ringle, M. H. (1982). Strategies for natural language processing. New York: Psychology Press.

Liu B (2010) Sentiment analysis and subjectivity, Handbook of natural language processing (2nd ed.) (2010).

Manav, B. (2017). Color-emotion associations, designing color schemes for urban environment-architectural settings. Color research and application, 42(5), 631-640.

Mikolov, T., Kombrink, S., Burget, L., Cernocky, J., & Khudanpur, S. (2011). Extensions of recurrent neural network language model. Acoustics, Speech and Signal Processing (ICASSP), 5528.

Moon, C. B., Kim, H., Lee, D. W., & Kim, B. M. (2013). Mood lighting system reflecting music mood. Color Research & Application, 40(2), 201.

Ou, L.-C., Luo, R., Wookcock, A., & Wright, A. (2004). Ou, L. C., Luo, M. R., Woodcock, A., & Wright, A. (2004). A study of colour emotion and colour preference. Part I: Colour emotions for single colours. Color Research & Application, 39(3), 232-240.

Pestian J, Nasrallah H, Matykiewicz P, Bennett A, Leenaars A (2010) Suicide note classification using natural language processing: A content analysis, Journal of Biomedical Informatics Insights, 3, 19-28.

Pizzi, D., Charles, F., Lugrin, J.-L., & Cavazza, M. (2007). Interactive Storytelling with Literary Feelings. Affective Computing and Intelligent Interaction, 630-641.

Talklife (2018) https://talklife.co/, retrieved May 2018.