A Decision-Level Approach to Multimodal Sentiment Analysis

Haithem Afli[1-2], Jason Burns[1] and Andy Way[1]

[1]ADAPT Centre
School of Computing
Dublin City University
Dublin, Ireland
{FirstName.LastName}@adaptcentre.ie
[2] Cork Institute of Technology
haithem.afli@cit.ie

Abstract. There has been near exponential increase in the use of images and video on various Social Media platforms in the last few years, in place of or in addition to the use of plain text. Automated sentiment analysis, at its core, is the capturing of human emotion by machine – the addition of image and video to social media output had made this already challenging task even greater. In this paper, we propose a multimodal, decision-level based approach to sentiment analysis (SA) of Twitter feeds. The solution proposed and outlined in this paper, combines the sentiment analysis scoring of not just text-based output but integrates SA scoring generated from analysis of image captions. For our experiments, we focused on politics and on two political topics (Trump/Brexit) that are generating a lot of discussion and debate on Twitter. We chose the political domain given the power that Social Media has on possibly influencing voters [1] and the 'strong' opinions that are expressed in this area.

Keywords: Multimodality, sentiment analysis, image captioning, tweets.

1 Introduction

The purpose of combining multi-modal data in order to produce better estimates for sentiment analysis, in other words improving the accuracy of estimating human opinion when expressed through a combination of multi-modal channels, is the goal of this work. To date, there has been limited research into multimodal sentimentality [3]. The growth in social media has been extraordinary and the power of certain social media platforms at the very least to disseminate information globally in a few minutes of the event happening is bound to have an effect on a large proportion of people's opinion. So, given the global reach and effect of social media, the ability to accurately monitor and analyse information, and the

[1] https://www.theguardian.com/technology/2016/jul/31/trash-talk-how-twitter-is-shaping-the-new-politics

sentiment scoring of that information, is crucial for a number of reasons. But the impact of media on world events is nothing new and as far back as 1971 research was being conducted on the influence of the media and world events on the stock market [10]. It is the informal structure (and use of 'non-canonical language' [5]) and content of social media output, and its sheer volume, that has led to several challenges in accurately processing and obtaining accurate sentiment analysis (SA) scores. Certain social media channels (such as Twitter) present their own challenges with either a limited character count, informal language usage, a lack of context and now an ever-increasing use of images and video.

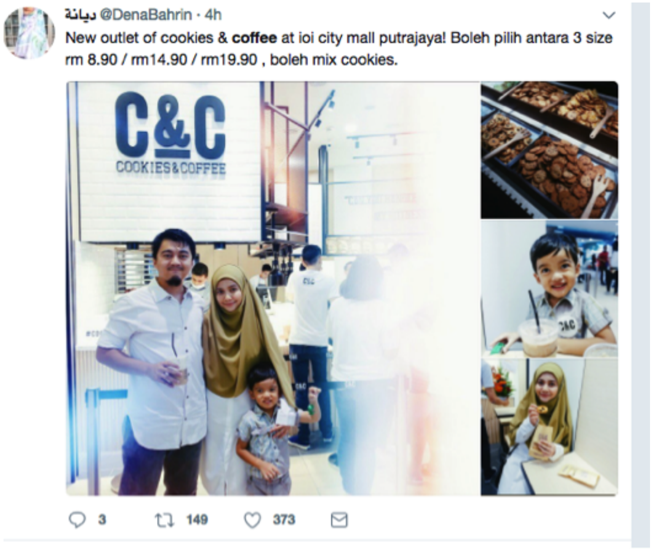
These challenges and need for obtaining accurate SA scores has led to a large amount of successful research on sentiment analysis of social media, largely focusing on some derivative of a text classification process [9, 2]. The analysis, for most part, makes a distinction between positive sentiment, negative sentiment, or neutral. But it is the ever-increasing use of multi-modal communication, namely images and video in addition to or as a replacement for text, that has led to a major challenge for any trying to conduct accurate SA on social media output. Up to recently, there is not a lot of work being carried out in addressing this 'multi-modal' sentiment problem where several different methods of communication are being used [11].

Twitter is one of the more common social media platforms used by people to express their opinions and emotions. The access available to harness and process tweets (for example, through the use of APIs) has led to some interesting studies from determining how certain users of Twitter can affect general opinion on the platform through to establishing the power of twitter on the outcome of US elections [12].

What is clear is that given the character restriction of Twitter to 140 characters and the power of images to convey a complex amount of information very quickly, there is a very large increase in the amount of images being included in tweets. What is also being established is that tweets are likely to be far more influential if they do contain an image. In 2012 for example, arguably the most popular tweet of the year was a simple image of Barack and Michelle Obama with little or no text included in many of the retweets.

However it is in the combining of information from multiple modes that is essential in improving the accuracy of sentiment analysis applications. Humans communicate in a multi-modal formats and therefore the analysis of text, audio and images/video is fundamental to improving our sentiment analysis accuracy levels [9]. The reason for this increased interest in multi-modal sentiment analysis is directly linked to the massive increase in the use of images and video through social media (Facebook and Youtube) in particular.

There are many practical applications of automated sentiment analysis in use today [7]. To just look at one example, as more and more product reviews are conducted through video, the market is more interested in learning opinions expressed through this medium than simply based on textual reviews. Therefore, marketing teams worldwide are looking at solutions where sentiment is more accurately tracked in relation to the use of their particular product or service. But the automation of SA is challenging when limiting your analysis to text alone. As can be seen in the image 2, the text is neutral in terms of sentiment, but it is clear from the images that the sentiment is positive simply by the smiling faces that are included with the tweet. This is a straight forward and simple, but powerful example of how images can enrich the sentiment that we can receive from tweets.



ديانة @DenaBahrin · 4h

New outlet of cookies & **coffee** at ioi city mall putrajaya! Boleh pilih antara 3 size rm 8.90 / rm14.90 / rm19.90 , boleh mix cookies.

Fig. 1. Sample Tweet of positive sentiment

For this work, we focused on Twitter as the social media channel of choice. The use of image and video, combined with text, is growing rapidly in Twitter and for good reason. A study by Buffer showed that including a simple image with your tweet was the most advantageous method of ensuring your tweet was more widely read. Studies have recently been carried out on the effect of adding mulitmedia to tweets within Sian Weibo (Chinese version of Twitter) and it is clear that adding media (particularly images) does increase the popularity and lifetime of that tweet [13].

The goal is to improve on the existing textual sentiment analysis solutions by including SA scores generated from image captions in order to determine if the overall SA score is improved by using this additional information. By using the very latest in image caption generation technology, we can unlock the information contained within the embedded image to give us a more accurate score on what sentiment is being expressed by a particular tweet. .

2 Related work

Based on the advances in image recognition, attention is now starting to focus on sentiment analysis of images and video, and this is too proving to be a major challenge. It is one thing to recognise objects in an image, possibly connect those objects and offer a context, but it is a major leap in understanding to try to determine an opinion or emotion from the image. The research is hampered by the lack of available datasets that properly annotate images with emotion tags, a suitable lexicon of images and associated sentiment categories. A few public datasets such as the International Affective Picture System (IAPS) [2] and the Geneva Affective Picture Database (GAPED) [4] being the only datasets available of any note that provide either ratings or annotations for emotion and sentiment.

The first step in sentiment analysis of images is to determine the content of images and form a sentence describing the content – it is substantially harder than just classifying images [8]. The human eye can gather an immense amount of information from glancing at an image, a capability that is proving to be very difficult for machines to replicate [6]. There are several reasons why conducting object recognition/image captioning on images and video is very difficult; not least because a very large dataset is needed to train models – in many instances a model needs a large amount of prior knowledge to compensate for the information that is missing [8].

3 Decision-Level Approach to Multimodal Sentiment Analysis

There were several phases to the project that are outlined in the following sections. In summary, searches were conducted on Twitter for specific popular topics (#Trump, #Brexit) and only tweets with images were stored and processed on IBM Bluemix before being analysed for sentiment using the SentiTweetWords [1] analysis tool and NeuralTalk2 [3].

3.1 Data Preparation and system architecture

We chose the IBM Bluemix Platform as a Service (PaaS) environment to collect and process the tweets, a cloud based service making it easier to access the

[2] `http://csea.phhp.ufl.edu/media/iapsmessage.html`

[3] https://github.com/karpathy/neuraltalk2

technology needed to process the tweets received. First we created a nodered.js app that would connect the Twitter API to our couchdb database (Cloudant) . Connected to the Twitter Decahose we searched for the hastags #Trump and #Brexit, and then stored the captured tweets in the Cloudant no-SQL database (based on the couchDB framework) as separate JSON objects. The JSON files were then copied, using Spark-based ETL tool, to the DashDB datawarehouse. Once the JSON objects were coped to DashDB it was easier to perform basic analysis to isolate the tweets with images included – the nominated tweets (and images) were then exported from Bluemix for processing.

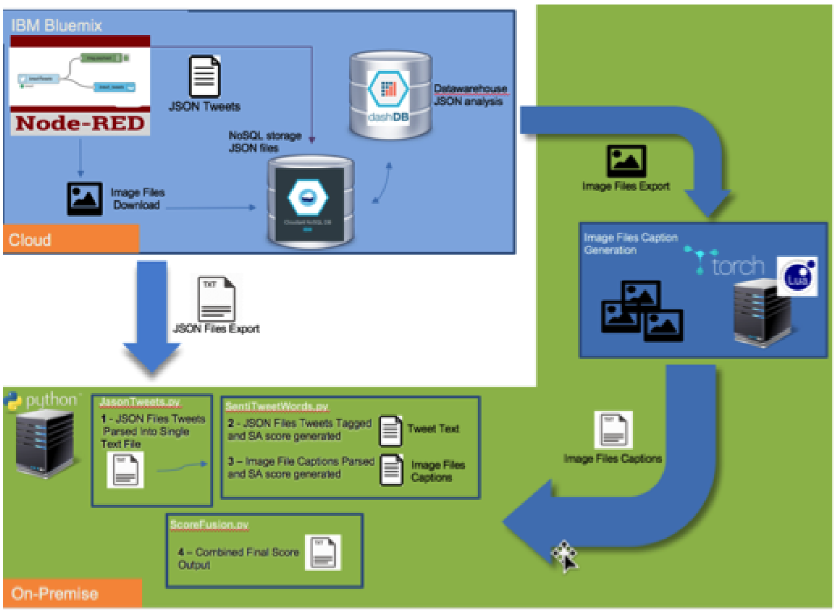


Fig. 2. Sample Tweet of positive sentiment

4 Experiments and results

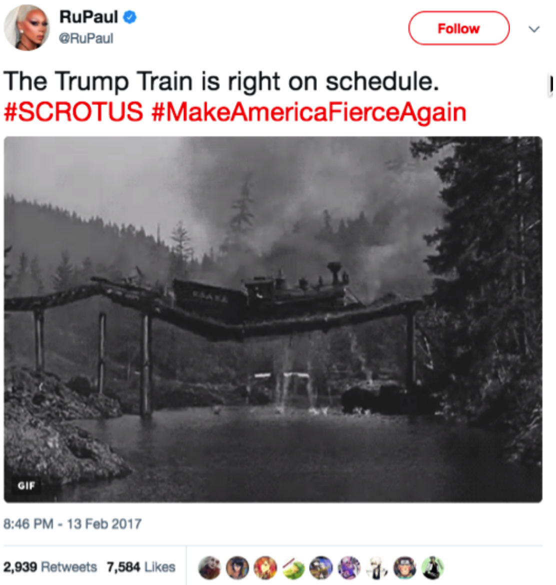
The collated scores for the tweet text and image captions were captured and averaged for analysis. In order to fully assess the sentiment analysis scoring for both text (SA_Text_Score) and image captions (SA_Image_Score) a longer scope of manual annotation would need to be undertaken –this is beyond the scope of this paper. However, taking a sample of 200 tweets, we were able to perform some elementary analysis of the scoring returned by the automated multimodal sentiment analysis application. Before addressing the results and scoring, there are some important with regard to our application and the general sentiment analysis environment:

- Image Recognition Maturity – In line with many leading-edge sub domains within the overall Artificial Intelligence/Machine Learning field, there is still a lot of scope for progression in the accuracy and completeness of image caption generation. Captions generated by NeuralTalk2 to are largely accurate but simple in content, therefore this affected the overall SA_Image_Score result.
- Informal Language Usage - Achieving an accurate SA score on tweets is difficult when sarcasm and very informal language is being used. This in turn affects and NLP application that is trying to determine sentiment, and therefore SA_Text_Score.
- Domain Choice – our domain of choice was Politics, and two popular topics in that area as of 2017: Brexit and Trump. As already discussed earlier in this document, we chose this domain because of the affect social media platforms can have on popular political opinion. However, it created some unique difficulties. For example, many prolific tweeters (and regular political commentators) expect that their political views are already known, and therefore provide little context to their tweets which makes it difficult to automate the SA process. Also, images can be used that are in direct conflict (in terms of sentiment) to the text of the tweet (sarcasm being used) and this can have an adverse affect on the overall averaged sentiment analysis score.

Both SA_Text_Score and SA_Image_Score had a sentiment analysis scoring range of between 0 to 1. Looking more closely at the actual results, focusing on the random 200 tweets, we observed the following:

- Close to 60% of tweets had SA_Text_Score and SA_Image_Score numbers that were very close in score (difference of <.19). There are several different interpertations to be taken from this - where the application can not determine a score it will default to neutral. It is therefore reassuring that both image and text correspond.
- <9% of scores were in contradiction to each other in terms of sentiment scoring – this is not necessarily a bad result. There are several possible reasons for this, some tweets contain a lot of textual information (relating to several different points) and therefore it is difficult to determine the sentiment accurately regardless of the image. As mentioned before, the use of sarcasm can confuse the application, therefore this is something that could be used as a trigger for detecting sarcasm
- When the image caption was accurate and the text of the tweet was clear to read, there was a very close correlation between SA_Text_Score and SA_Image_Score. ˜10%. Given the challenges that are in place for our application this again was a positive result highlighting the fact that when some obstacles were removed, the application was accurate in predicting an SA score.

As can be seen in the example of figure 3, an image can either confirm opinion or present an alternative opinion, especially where the use of sarcasm is being employed.



RuPaul
@RuPaul

The Trump Train is right on schedule. #SCROTUS #MakeAmericaFierceAgain

8:46 PM - 13 Feb 2017

2,939 Retweets 7,584 Likes

Fig. 3. Sample Tweet of positive sentiment

5 Conclusion and Future Work

We have presented a multimodal sentiment analysis that builds on the existing work done on textual sentiment analysis and image captioning. By combining the different modes, we have shown that you can enrich the sentiment analysis score of the tweet by including the image caption in the sentiment scoring. There are some difficult challenges that need to be overcome, but we have presented some possible improvements that might go some way to alleviating the factors that affect accurate automated SA scoring. Twitter is a fast-moving and prolific medium for opinion expression, and focusing on the political domain served to highlight clearly the range, diversity and polarity of opinions that are expressed and shared on the Twitter platform. What is clear from our research, is that images and video are only going to increase in usage and it is no longer viable for a social media sentiment analysis application to ignore those media.

Acknowledgements

References

1. Haithem Afli, Sorcha McGuire, and Andy Way. Sentiment translation for low resourced languages: Experiments on irish general election tweets. In *Proceedings of the 18th International Conference on Computational Linguistics and Intelligent Text Processing*, Budapest, Hungary, 2017.

2. Claudio Baecchi, Tiberio Uricchio, Marco Bertini, and Alberto Del Bimbo. A multimodal feature learning approach for sentiment analysis of social network multimedia. *Multimedia Tools Appl.*, 75(5):2507–2525, March 2016.
3. Erik Cambria, Bjorn Schuller, Yunqing Xia, and Catherine Havasi. New avenues in opinion mining and sentiment analysis. *IEEE Intelligent Systems*, 28(2):15–21, March 2013.
4. Elise S. Dan-Glauser and Klaus R. Scherer. The geneva affective picture database (gaped): a new 730-picture database focusing on valence and normative significance. *Behavior Research Methods*, 43(2):468, Mar 2011.
5. Jacob Eisenstein. What to do about bad language on the internet. In *Proceedings of NAACL-HLT 2013*, pages 359–369, Atlanta, Georgia, 2013.
6. Andrej Karpathy and Li Fei-Fei. Deep visual-semantic alignments for generating image descriptions. *IEEE Trans. Pattern Anal. Mach. Intell.*, 39(4):664–676, April 2017.
7. Siavash Kazemian, Shunan Zhao, and Gerald Penn. Evaluating sentiment analysis evaluation: A case study in securities trading. In *Proceedings of the 5th Workshop on Computational Approaches to Subjectivity, Sentiment and Social Media Analysis*, pages 119–127, Baltimore, Maryland, USA, 2014.
8. Alex Krizhevsky, Ilya Sutskever, and Geoffrey E. Hinton. Imagenet classification with deep convolutional neural networks. *Commun. ACM*, 60(6):84–90, May 2017.
9. Louis-Philippe Morency, Rada Mihalcea, and Payal Doshi. Towards multimodal sentiment analysis: Harvesting opinions from the web. In *Proceedings of the 13th International Conference on Multimodal Interfaces*, ICMI '11, pages 169–176, New York, NY, USA, 2011. ACM.
10. Robert P. Schumaker. An analysis of verbs in financial news articles and their impact on stock price. In *Proceedings of the NAACL HLT 2010 Workshop on Computational Linguistics in a World of Social Media*, WSA '10, pages 3–4, Stroudsburg, PA, USA, 2010. Association for Computational Linguistics.
11. Rajiv Ratn Shah. Multimodal-based multimedia analysis, retrieval, and services in support of social media applications. In *Proceedings of the 2016 ACM on Multimedia Conference*, MM '16, pages 1425–1429, New York, NY, USA, 2016. ACM.
12. Hao Wang, Dogan Can, Abe Kazemzadeh, François Bar, and Shrikanth Narayanan. A system for real-time twitter sentiment analysis of 2012 u.s. presidential election cycle. In *Proceedings of the ACL 2012 System Demonstrations*, ACL '12, pages 115–120, Stroudsburg, PA, USA, 2012. Association for Computational Linguistics.
13. Xun Zhao, Feida Zhu, Weining Qian, and Aoying Zhou. Impact of multimedia in sina weibo: Popularity and life span. In *Semantic Web and Web Science - 6th Chinese Semantic Web Symposium and 1st Chinese Web Science Conference, CSWS 2012, Shenzhen, China, November 28-30, 2012.*, pages 55–65, 2012.