Memotion Analysis : Automatic processing of Internet Memes

Subalakshmi Shanthosi S (186001008)
ME CSE, Semester 3

Dr. Aravindan Chandrabose
Supervisor

Project Review: 1 (14 August 2019)
Department of Computer Science and Engineering
SSN College of Engineering

---

1 Introduction

In recent years, a huge amount of User Generated Content(UGC) online of various modalities like text,images and videos are accumulated on the web. NLP and Computer Vision communities often leverage only one prominent modality(text) in isolation to study social media.But,computational processing of internet meme needs a hybrid approach.This research work aims at building an automatic processing of Internet memes.

2 Motivation

Detection of offensive content on social media is an ongoing struggle due to the following reasons:
1. More number of emerging social websites creating voluminous data.
Ex:Facebook,Flickr,Twitter,PinInterest etc.

2. Prevalance of hate speech in social media is profound and it has become a great societal responsibility for the government and various social media companies to plan well ahead for mitigation and eventual prevention.

3. Multimodal sentiment analysis is still in initial stage as it is less explored by many.

2.1 Challenges in Memotion Analysis

: 1. A meme is uniquely multimodal as it has both visual and textual descriptors.
2. Most of the internet meme show replication in content by having same meme template with different sentences conveying same semantic meaning which are to be considered as duplicated in order to ensure better accuracy.

3. Challenges in doing OCR compound when the number of potential fonts, languages, lexicons and other special characters increase.
Tesseract OCR isn't capable of handling italics and other special character encodings.
4.Defining correct visual descriptors which are to be evaluated in order to access the emotion is still a unanswered question.
5. Emotion Semantic Image Retrieval(ESIR) with affective gap making the low-level image features extrapolate the high-level semantics. High-level features are nothing but textual descriptors like concept name, keywords etc. Low-level features are qualitative measures of images like colour, texture, shape,spatial layout etc.

3 Problem statement

Memes typically induce humor and strive to be relatable. Some memes are directly humorous whereas others go for sarcastic dig at daily life events.Three subtasks varied by the degree of exploration are as follows:

1. **Task A - Sentiment Classification**: Given an Internet meme, the first task is to classify it as positive or negative meme. We presume that a meme is not neutral.

2. **Task B- Humor Classification**: Given an Internet meme, the system has to identify the type of humor expressed. The categories are sarcastic, humorous, and offensive meme. If a meme does not fall under any of these categories, then it is marked as a other meme.

3. **Task C- Scales of Semantic Classes**: The third task is to quantify the extent to which a particular effect is being expressed(refer Table 1).

8K human annotated Internet memes labelled with semantic dimensions namely sentiment, and type of humor that is, sarcastic, humorous, or offensive.

The humor types are further quantified on a likert scale as in Table 1. The dataset will also contain the extracted captions/texts from the memes.

Table 1: Semantic classes for the Memotion Analysis

	sarcastic	**humorous**	**offensive**
not (0)	😐	😐	😐
slightly (1)	😋	😄	😒
mildly (2)	😉	😆	😠
very (3)	😜	😂	😤

4 Literature survey

4.1 Image-based memes as sentiment predictors[1]

Opinion extraction by finding correlation between the implied semantic meaning of image-based memes and textual discussions in social media.Additionaly, trends in the use of popular memes are also identified. Visual predictors like colour,depth and shapes are used for image categorisation.
Dataset Used:
Facebook:Public posts and discussions.
10 discussion forums are taken into account with 997 unique comments and 103 memes.
A total of 27,260 words are existing.
Also for dynamic dataset Random discussions pages containing memes are taken for processing.

Category	Description
Word Analysis	Based on each words in a discussion. Sentiment score calculated by using matching algorithm to sentiment dictionary.
Phrase Analysis	Semantria Sentiment Library to assign sentiment scores for each phrase in the discussion. Can also use phrase intensifiers like always or never.
Comment Analysis	Averaged phrase analysis score for all the comments in one discussion thread.
Discussion Analysis	Based on both word level and phrase level analysis. Synonyms dictionary is constructed for ”keyterm” and scores calculated by comparing to find presence of that word.

Table 1: Textual Analysis of Facebook Discussions

Category	Description
Meme Textual Descriptions	84% memes are textual hence OCR extracted texts are used for sentiment analysis. Visual meme descriptors are found by finding appropriate action word i.e) closest meaning word which describes that action.
Meme Category	Memes provided with categorical tag. Tags:35,unique and descriptive entities Ex:popcorn,teeth
Meme Popularity	Frequency analysis of most common meme category which where used.In addition to it existence of particular tag is found on discussion forum as well.

Table 2: Meme Analysis

Shortcomings:

•Misspelled words have NEUTRAL sentiment.

• Contradicting word-level and phrase level analysis.

Future Work:

• Sentiment dictionaries required addition of colloquial words or sensitive contextual words like ”banned” which have negative impact in Facebook but not found in sentiment dictionaries.

4.2 Memetic Engineering to Classify Twitter Lingo[2]

Sentiment Analysis on textual memes is the main focus on this research work.Combination of Machine Learning and Image Processing Techniques to leverage maximal computation of meme emotion.

Datasets:
1.Benchmark Dataset:CSV file containing 1524 tweets labelled as:

- POSITIVE(1)
- NEGATIVE(-1)
- NEUTRAL(0).

2.Dynamic Dataset using Twitter API:
Tweets gathered for a keyword given by the user.User login credentials are hashed and kept for finetuning dataset file inorder to login with those credentials to search for that keyword.

Machine Learning Techniques:

- Multinomial Naïve Bayes Algorithm (Performant).
- K-Nearest Neighbours(KNN) (More Performant).
- Support Vector Machine(SVM) (Less Performant).
- Logistic Regression(LR) (Moderate Performant).

Advantages:
Efficient sentiment evaluator which can be easily integrated to enterprise applications. KNN applicability even on small dataset.
Shortcomings:
Performance of OCR determines the accuracy of the system. Visual descriptors are not given much attention for emotion evaluation.
Future Work:
Making the system multi-lingual and scaleble to volumnious data.

4.3 Meme Opinion Categorization by Using Optical Character Recognition (OCR) and Naïve Bayes Algorithm[3]

This research work works on image-textual memes wherein the textual descriptor like comment/title phrases are extracted using Tesseract OCR.
Accuracy of OCR Tesseract is 75%.
Dataset:
Meme Collection which specifically deals with viral memes about indonesian government.
Total Number of memes taken:100.
Randomly split in the ratio of 70% and 30% for training and testing respectively.
Experimental Results: Naïve Bayes algorithm shows fairly good accuracy of 75%.
Vmap tendency is calculated to find P_{max} of all categories of documents tested.
Shortcomings:
OCR less accurate:Inability to recognise italics.
OCR incapable of recognising special characters and emoticons.
Future Work:
KNN,Neural network can be employed instead on Naïve Bayes as KNN could work on relatively smaller and low quality memes. n-gram tokenisation for enhanced preprocessing.

4.4 Rosetta: Large Scale System for Text Detection and Recognition in Images[4]

Rosetta is Facebooks scalable OCR system.
Rosetta aims at becoming most scalable and robust OCR system capable of handling millions of requests per day.
Two stage process:
Rosetta performs OCR in two independent steps:

1. Detection of rectangular regions in the image potentially containing text.

2. Text Recognition is performed on each detected regions using a CNN and then transcribe the word in the region.

Text Detection Model:
Approach based on Faster-RCNN a state-of-art object detection network.Faster-RCNN does detection and recognition simultaneously.
Supervised end-to-end process where each region's learning classifiers found (say k),which are sorted by their confidence score and non-maximum supression(NMS).
Alteration in Faster-RCNN - Replacement of ResNet convolutional body with a ShuffleNet-based architecture for improved efficiency.
The ShuffleNet convolutional body is pre-trained using ImageNet dataset.

Text Recognition Model:

1. **Character sequence encoding model(CHAR)**:
   Assumption:
   • All images are of same size.
   • Only k of it's characters are recognisable from a word regardless of the word length.
   Working:
   The body of the CHAR model consists of a series of convolutions followed by k independent multiclass classification heads, each of which predicts the character of the alphabet (including the NULL character) at each position. During training, one jointly learns the convolutional body and the k different classifiers.

2. **Fully Conventional Model- CTC**:
   Known as CTC model because it uses sequence-to-sequence CTC loss during training. After convolutional body which is ResNet-18 , the last Convolutional layer predicts the most likely character at every image position of the input word.
   Example: LEARNING, the model might produce the sequence of characters "L-EE-A-RR-N-I-NN-G", which includes blanks and duplicates.
   Approach:
   • Initialize the weights of the model body with the trained weights of the CHAR model, and then finetune those weights while simultaneously learning the last convolutional layer from scratch.
   •The second approach was based on curriculum learning , i.e., starting with a simpler problem and increasing the difficulty as the model improves.Starting with word length 3 and gradually increasing them at every epoch.

Datasets:
1.COCO-Text which contains:

- 63,000 images.
- 145,000 text instances.

2.Large synthetic dataset for covering maximum usecases as COCO-Text doesn't match the data-distribution of images uploaded to Facebook.

- 400k images for training.
- 50k images for testing.

Experimental Results:
Error rate of 37% still recoverable by changing single character.
Finetuning with manually annotated corpus increased accuracy by 48.06%
Random jitter introduced for data augmentation.
Advantages:
Robust and accurate OCR capable of processing millions of images per day.
Faster search from TAO for recognised text.
Adaptive character based recognition.
Limitation:
Resolution of image increases inference time.
Future Work:
Case sensitive labelling affected performance.

4.5 An image-text consistency driven multimodal sentiment analysis approach for social media[5]

Visual feature detection using Local Binary Pattern. Textual Feature extraction: Continuous bag-of-words or skip-gram. Image-text similarity found.
Datasets:
Visual Sentiment Ontology:
Total number of images: 603.
Topics:16.
Training and Testing:
Training Dataset :400 images.
Test Dataset: 157 images.

Results: F-score:
Positive category:0.87 Negative category:0.89
Adaptive merging of textual features with State-of-art SentiBank for improved accuracy.
Advantages:
Superior performance in Flickr benchmark dataset.
Use of AdjectiveNounPair(ANP): Converting neutral noun into Strong sentiment word.
Shortcomings and Future Work:
SVM modelling unrelated and related data sensitive to outliers.
ANP difficult due to its abstract nature and high variability

5 Existing system

- The existing approaches have used meme- discussion text correlation for finding concordance or using Pretrained CNN model for extracting top image content descriptors closet to the context. [1][5].
- Rosetta method of Extraction and pre-processing for additional data set creation can be done for multilingual memotion analysis[4]
- KNN or Naïve Bayes classifier can be used for maximising F-score.[3][2]

6 Proposed system

- Pre-trained CNN to work on either image or OCR extracted text classifier based on the image-text correlation.
- Increasing working dataset size by using Rosetta Facebook's OCR system to make the system more generic by handling multiple languages.
- KNN approach or Naïve Bayes classifier for opinion extraction to categorise as positive or negative.
- Using ResNet for faster computation.

References

[1] Jean H. French, *Image-based Memes as Sentiment Predictors.* International Conference on Information Society (i-Society 2017).

[2] S. Priyashree, N. Shivani, D. K. Vigneshwar,S. Karthika, *International Conference on Computational Intelligence in Data Science(ICCIDS)*, 2017.

[3] Amalia Amalia , Arner Sharif , Fikri Haisar , Dani Gunawan , Benny B Nasution ,*Meme Opinion Categorization by Using Optical Character Recognition (OCR) and Naïve Bayes Algorithm.* Third International Conference on Informatics and Computing (ICIC), 2018.

[4] Bogdan Lazarescu, Christo Lolov , Silvia Sapora, *Rosetta: Large scale system for text detection and recognition in images.*, KDD'18 Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining,Pages 71-79,2018.

[5] Ziyuan Zhao , Huiying Zhu , Zehao Xue and Zhao Liu et al.,*An image-text consistency driven multimodal sentiment analysis approach for social media* , Information Processing and Management Vol 56-Issue 6,2019.