**Clustering Algorithms in Relation**

**to Word Analysis to Understand Context –**

Jephone Torre

https://www.sciencedirect.com/science/article/abs/pii/S0306457319306934

Abstract

Detecting sentiments in natural language is tricky even for humans, making its automated detection more complicated. This research proffers a hybrid deep learning model for fine-grained sentiment prediction in real-time multimodal data. It reinforces the strengths of deep learning nets in combination to machine learning to deal with two specific semiotic systems, namely the textual (written text) and visual (still images) and their combination within the online content using decision level multimodal fusion. The proposed contextual ConvNet-SVMBoVW model, has four modules, namely, the discretization, text analytics, image analytics, and decision module. The input to the model is multimodal text, m ε {text, image, info-graphic}. The discretization module uses Google Lens to separate the text from the image, which is then processed as discrete entities and sent to the respective text analytics and image analytics modules. Text analytics module determines the sentiment using a hybrid of a convolution [neural network](https://www.sciencedirect.com/topics/social-sciences/neural-network) (ConvNet) enriched with the contextual semantics of SentiCircle. An aggregation scheme is introduced to compute the hybrid polarity. A support vector machine (SVM) classifier trained using bag-of-visual-words (BoVW) for predicting the visual content sentiment. A Boolean decision module with a logical OR operation is augmented to the architecture which validates and categorizes the output on the basis of five fine-grained sentiment categories (truth values), namely ‘highly positive,’ ‘positive,’ ‘neutral,’ ‘negative’ and ‘highly negative.’ The accuracy achieved by the proposed model is nearly 91% which is an improvement over the accuracy obtained by the text and image modules individually.

Introduction

Social media has enabled mobilization of information where users can post and share all kinds of multimodal text in the social setting without much knowledge about the Web's client-server architecture and network topology. Eliminating communication and demographical barriers it serves as a communication channel, social listening, and feedback tool for stakeholder engagement and cooperation. Nevertheless, organizations and big businesses are keen to develop applications that support automated text analytics, deriving meaningful information from the high-diversity, multimodal data is a crucial aspect.

Sentiment Analysis (Pang & Lee, 2008; Cambria, 2016) is touted as the key to unlock big data in the social setting for practical data-driven decision making. It is defined as a generic text classification task which indispensably relies on the understanding of the human language and emotions expressed in the social media post. There are different ways to model human emotion, the affective spectrum, and the subjectivity. The aesthetics of sentiments in social psychology lies within the universal field of mind, spirit, and body with the conscious level of emotional processing (Fig. 1). Emotions, feelings, and core affect, define the affective phenomena where core affect is an outward expression of our feelings and emotion. Though both emotions and feelings are often used interchangeably, the two are quite distinct. Emotions are bodily, instinctive, and quantifiable. They can be measured with the help of blood flow, heartbeat, brain activity, facial expressions, and body movements.

On the other hand, feelings are created by the senses, often fueled by a mix of emotions, and last for longer than emotions. For example, ‘satisfied’ and ‘grateful’ are two sample feelings created by the emotion ‘love.’ Emotions may strengthen and define an attitude which describes the way humans act or react to people or situation. Simultaneously, emotions can further trigger the mood (hours or days), subsequently prompting a sentiment which can persist indefinitely. Further, sentiments about a particular subject matter (topic, object, event, person or situation) define an opinion or view. It defines an informational sentiment characterized by a quintuple <entity, aspect, sentiment, holder, time>, where, entity is the object/target entity; aspect is the feature of the entity, sentiment is the polarity or rating, holder is the opinion holder and time is the time of opinion expression (Liu, 2015).

Pertinent literature shows sufficient evidence of methods, systems, and applications within the domain (Pak & Paroubek, 2010; Cambria, Schuller, Xia & Havasi, 2013). The findings and learning from relevant studies embody two primary techniques for analyzing the sentiment, namely the machine-learning enabled techniques and the lexicon-based techniques on user-generated online content (Pang, Lee & Vaithyanathan, 2002; Kumar & Jaiswal, 2019).

The language and linguistic tone of user-generated content are informal and indistinct. Recent observations exemplify an array of language constructs and usage styles which include the use of emblematic language markers such as punctuation (super!!!!!!), emoji (), micro-text (Satapathy, Guerreiro, Chaturvedi & Cambria, 2017) and multilingual typing. All this increases the complexity of computational linguistics to analyze social media content. Further, analyzing explicit and clear sentiment is challenging owing to language constructs which may intensify or flip the polarity within the posts. For example, the tweet, “He is really good at cheating” conveys sarcasm which is challenging to understand without contextual cues. That is, without context, this sample tweet is classified as positive because of the presence of the term ‘good’ in it. It is only when the context of the word ‘good’ is taken into consideration; it is categorized as negative or unfavorable since the word ‘cheating’ is a negative polarity word. Thus, it is imperative to comprehend additional cues from users’ linguistic input that are aware of ‘context’ which aid right interpretation. However, understanding context is one of the most challenging aspects of content moderation. Besides, contextual assistance has been studied across pertinent literature; its effectiveness in sentiment analysis needs further validation.

Also, as more recently, memes (viral image, video or verbal expression for mimicry or humorous purposes), animated GIFs (Graphics Interchange Format which combines multiple images or “frames” in a single file to convey motion), typo-graphic (artistic way of text representation), info-graphic (text embedded along with an image) visual content, and edited videos dominate the social feeds. Further, the intra-modal modeling and inter-modal interactions between the textual, visual, and acoustic components add to the linguistic challenges. A text could perhaps be well-defined as multimodal when it combines two or more semiotic systems to create meaning (Fig. 2). Typically, semiotics is an investigation into how meaning is created and how meaning is communicated. The semiotic systems can be categorized as follows (The New London Group 2000):

•Linguistics: vocabulary, structure, grammar of oral/written language

•Visual: color, vectors, and viewpoint in still and moving images

•Aural: volume, pitch, and rhythm of music and sound effects

•Gestural: movement, facial expression, and body language

•Spatial: proximity, direction, the position of layout, organization of objects in space

Interestingly, the multimodal social text is estimated to be 90% unstructured making it crucial to tap and analyze information using contemporary tools. There is extensive use of multimodal social media platforms which allow expression of opinion using videos (for instance: YouTube, Vimeo, VideoLectures), images (e.g., Flickr, Picasa, Facebook) and audios (e.g., podcasts). The machines now need to extend the cognitive capabilities to interpret, comprehend, and learn features over multiple modalities of data acquired from different media platforms. Thus, the research on sentiment analysis warrants a new line of inquiry to understand how representation learning and shared representation between different modalities and the heterogeneity of the multimodal data challenges the performance of models.

Multimodal sentiment analysis intends to apprehend varied sentiment evidence from the data with different modalities (a combination of text and audio-visual inputs). Pertinent literature studies report multimodal fusion as a task of avidly processing this mix modality of textual, audio, and visual features to facilitate improved understanding of opinions in user-generated content. Technically, multimodal fusion is the concept of integrating information from multiple modalities with the goal of predicting an outcome measure: a class (e.g., happy vs. sad) through classification, or a continuous value (e.g., the positivity of sentiment) through the regression (Baltrusaitis, Ahuja & Morency, 2017). Multimodal fusion techniques can be broadly categorized into two types as shown in Table 1:

Despite recent advances within the domain of multimodal fusion for sentiment analysis, three key challenges persist (Baltrusaitis, Ahuja & Morency, 2017; Majumder, Gelbukh, Hazarika & Cambria, 2018; Poria S Majumder et al., 2018):

•Difficulty in building models that exploit both supplementary and complementary information

•Different modalities may carry conflicting information

•Difficulty in efficiently capturing the intra-modality dynamics

Therefore, to ensure a reliable decision making (classification), accuracy of polarity classification depends on improved quality of feature vectors (both unimodal and multimodal) and the learning model. Motivated by this, we put forward a context-aware decision level fusion model for multimodal sentiment analysis in multimodal text, m, where m ε {text, image, info-graphic}. Deep learning architectures have proven capabilities for extrapolating new features from a limited set of features contained within a training set, without human intervention and without the need to label everything. These have given excellent results in comparison to conventional machine learning techniques for various natural language processing task (Zhao et al., 2019). At the same time, contextual clues can help detect fine-grained sentiment from text by resolving the ambiguity of meaning and improving the generic polarity classification. Based on these capabilities, the proposed contextual ConvNet-SVMBoVW model is a hybrid of ConvNet enriched with the contextual semantics of the SentiCircle (Saif, Fernandez, He & Alani, 2014) approach for predicting the textual sentiment and a bag-of-visual-words (BoVW) (Tirilly, Claveau & Gros, 2008) trained support vector machine (SVM) classifier for predicting the visual content sentiment. The info-graphic content is discretized by separating text from the image using Google Lens of Google Photos App.1 The processing of textual and visual components is carried out using hybrid architecture. A Boolean system with a logical OR operation is augmented to the architecture which validates and categorizes the output on the basis of five fine-grained sentiment categories (truth values), namely ‘highly positive,’ ‘positive,’ ‘neutral,’ ‘negative’ and ‘highly negative.’ This unifying model thus considers modalities of content and processes each modality type using a concord of deep learning and machine learning techniques for efficient decision support for sentiment analysis. The generic architectural workflow of the proposed model is given in Fig. 3

Thus, the key contributions of the work are:

•Individual, as well as mix of textual and visual semiotic modalities of social data, namely, textual, visual and info-graphic (text embedded along with an image), are taken into account.

•As analyzing explicit and clear sentiment in written text is challenging owing to language constructs which may intensify or flip the polarity within the posts. We propose the use of additional cues from users’ linguistic input that is aware of ‘context’ and aids right interpretation. A context enriched deep learning model for textual (written text) sentiment analysis is put forward. The model uses a convolution neural network (ConvNet) enhanced with the contextual scoring mechanism of SentiCircle.

•Multi-Class sentiment classification is proposed with polarity categorized into five fine-grained levels, namely, highly positive, positive, neutral, negative and highly negative.

The rest of the paper is organized as follows: The next section, Section 2 describes the related work followed by a detailed illustration of the proposed contextual ConvNet-SVMBoVW model for fine-grained sentiment analysis in multimodal online content in Section 3. Section 4 gives the results, and finally Section 5 concludes the research conducted.

**The proposed hybrid contextual ConvNet-SVMBOVW model:**

The proposed deep classification model reinforces the strengths of deep-learning nets in combination with machine learning to deal with different modalities of data in online social media content. The proposed Hybrid Contextual ConvNet-SVMBoVW model consists of four modules, namely, discretization module, text analytics module, image analytics module, and decision module (Fig. 4).

**Results and discussions**

The dataset prepared for experiments contains 8000 comments and posts (text, image, and info-graphic) prepared using the #CWC2019 on two social media sites Instagram and Twitter. The modalities within the dataset were 55% text, 15% images, and 25% info-graphic (Fig. 15).

**Conclusion:**

As the opportunities to analyze, model and discover knowledge from the social web applications/services are no more restricted to the text-based linguistic data but extend to the partially unknown complex structures of image, audio, and videos, novel challenges transpire to leverage this high-diversity multimodal data. This research proposed a hybrid model for real-time sentiment analysis in mix of text and image modality (info-graphic).

**Related works:**

-“Sentiment Analysis” (Dave, Lawrence & Pennock D, 2003) on all modalities (text, image, video, audio) of social data has been reported in the literature

A. Kumar et al.

[Empirical study of Twitter and Tumblr for sentiment analysis using soft computing techniques](https://www.sciencedirect.com/science/article/pii/B9780128113110000016)

N. Majumder et al.

[Multimodal sentiment analysis using hierarchical fusion with context modeling](https://www.sciencedirect.com/science/article/pii/S0950705118303897)

Knowledge-Based System (2018)

W. Medhat et al.

[Sentiment analysis algorithms and applications: A survey](https://www.sciencedirect.com/science/article/pii/S2090447914000550)

Ain Shams Engineering Journal (2014)

S. Poria et al.

[A review of affective computing: From unimodal analysis to multimodal fusion](https://www.sciencedirect.com/science/article/pii/S1566253517300738)

Information Fusion (2017)

S. Poria et al.

[Fusing audio, visual and textual clues for sentiment analysis from multimodal content](https://www.sciencedirect.com/science/article/pii/S0925231215011297)

Neurocomputing (2016)

S. Poria et al.

[Ensemble application of convolutional neural networks and multiple kernel learning for multimodal sentiment analysis](https://www.sciencedirect.com/science/article/pii/S0925231217302023)

Neurocomputing (2017)

K. Ravi et al.

[A survey on opinion mining and sentiment analysis: Tasks, approaches and applications](https://www.sciencedirect.com/science/article/pii/S0950705115002336)

Knowledge-Based Systems (2015)

M. Soleymani et al.

[A survey of multimodal sentiment analysis](https://www.sciencedirect.com/science/article/pii/S0262885617301191)

Image and Vision Computing (2017)

Alper Kursat Uysal et al.

[The impact of preprocessing on text classification](https://www.sciencedirect.com/science/article/pii/S0306457313000964)

Information Processing & Management (2014)

Rui Zhao

[Deep learning and its applications to machine health monitoring](https://www.sciencedirect.com/science/article/pii/S0888327018303108)

Mechanical Systems and Signal Processing (2019)