International Journal of Business Analytics & Intelligence
9 (1 & 2) 2021, 47-52
http://publishingindia.com/ijbai/

Visual Sentiment Exploration of Customer Emotions using Image Analytics

Riktesh Srivastava*

Abstract

Sentiment analysis is one of the vital areas to evaluate customer emotions. The growing prominence of sentiment analysis is because of social network platforms, which companies use for 360-degree consumer feedback. Companies use sentiment analysis as an automated process of recognising positive and negative emotions in online text data. By examining sentiments in social media comments and reviews, businesses can better understand how customers feel about their brands and products. Visual sentiment analysis aims to understand how images affect people in terms of evoked emotions. Companies are exposed to consumers' images on social media by consumers, and they need image analytics for social listening and response. In this paper, we took 21 random pictures from social media to identify the visual sentiment analysis. We use the image embedding algorithm in Inception V3, and Liu Hu and Ekman Algorithm to calculate the outcomes' polarity. Further, we used the machine learning classification algorithm to identify which model does the accurate classification of evoked emotions as happy and sad. Classification algorithms are based on the 2,048 features generated by the Inception V3 algorithm, and evoked emotions are classified accordingly.

Keywords: Inception V3, Liu Hu Algorithm, Ekman Polarity, Image Analytics, Picture Polarity, Naïve Bayes, Support Vector Machine, Neural Networks, Random Forest

Introduction

The advent of social media has provided vast amounts of information that is potentially valuable for social listening. Images taken randomly and then posted on social media give ample opportunities to identify these image sentiments (Basnyat, Anam, Singh, Gangopadhyay, & Roy, 2017). Enough research was conducted to assess emotions based on text; finding feelings from images has been chiefly ignored (Wang & Li, 2015). Researchers were performing analysis from image contents for various other fields, such as forensic analysis (Castiglione, Cattaneo & De Santis, 2011), health analysis (Garimella, Alfayad & Weber, 2016), topological image analysis (Almgren, Kim & Lee, 2017), disaster response analysis (Alam, Imran & Ofli, 2017), forgery identification (Maigrot, Kijak & Claveau, 2018), and food trend analysis (Amato et al., 2017), among others. With the rapid spread of social media, the analysis of sentiments from images has attracted research focus with broad prospects (Islam & Zhang, 2016; Ji, Cao, Zhou & Chen, 2016).

Earlier, social media posts (on Facebook, Twitter) were only in the form of texts, which extended to images and videos, with some platforms dedicated to posting pictures and videos (Instagram, Snapchat). There were a few remarkable observations:

- Tweets with images are 34% more likely to get retweeted than tweets with no pictures (Webster, 2015).
- Instagram users upload 55 million photos a day to the site (Bakhshi, Shamma & Gilbert, 2014).
- Photos on Facebook generate 53% more likes than the average post without photos (Corliss, 2019).

These shifts indicate that image-based posts possess a more significant social media presence than text alone (Pittman & Reich, 2016). With new technologies and smartphones, social media's future is more sensory-rich (Appel, Grewal, Hadi & Stephen, 2019).

In this paper, we do image embedding using the Inception V3 algorithm and comparative classifications using machine learning classification algorithms. The images are of the consumers who bought products online and posted their reactions, or an advertisement by companies for various products. As a sample, 21 pictures were

* City University College of Ajman, Ajman, UAE. Email: riktesh.srivastava@gmail.com

downloaded to identify evoked emotions. The paper is divided into 4 sections. Section 2 explains the proposed image analytics model, whereas section 3 gives the classification algorithms' outcomes using 4 parameters. Section 4 gives the conclusion and recommendations.

Image Analytics Model

The proposed model for image analytics is divided into 2 steps, as mentioned in Fig. 1. The first step in the model is image embedding, which takes the images as input and uses Inception V3, along with sentiment algorithm, to identify the visual sentiments. Around 21 different images used for the experiment were tested with Inception V3 and sentiment algorithms. The second part of the proposed model tests the efficiency between the four classification machine learning algorithms. The outcomes were evaluated using classification accuracy (CA), precision, recall, and F1 parameters.

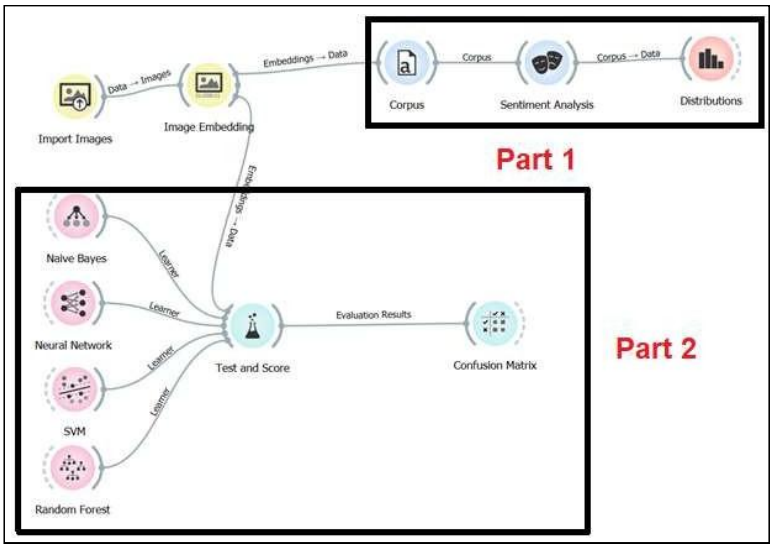


Fig. 1: Image Analytics of Social Media

Part 1: Inception V3 and Sentiment Algorithm

Developed by Google and trained in ImageNet, Inception V3 uses the activations from the model's penultimate layers, characterising an image with vectors (Szegedy, Vanhoucke, Ioffe, Shlens & Wojna, 2015). Inception V3 uses fewer parameters, with 42 deep learning networks to identify the images (Tsang, 2015).

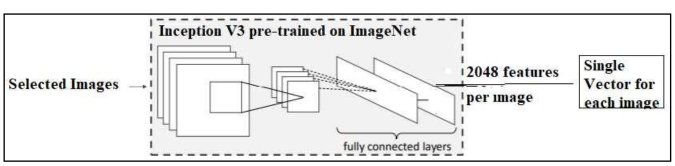


Fig. 2: Image Embedding using Inception V3

Around 21 images placed in Inception V3 algorithm are pre-trained on ImageNet, and the last fully connected layer activations are extracted as large deep spectrum feature vectors. It results in 2,048 features for each image, which is bagged into a single feature vector per image.

The steps can be represented mathematically by the following equation.

$$\sum_{i=1}^{21}\prod_{j=1}^{2048} Image_{ij} = \sum_{i=1}^{21}\left[\prod_{j=1}^{2048}\left[\begin{bmatrix}(f_{11})\\(f_{22})\\ \vdots\\ \vdots\\(f_{Nj})\end{bmatrix}\right]\right] = \sum_{i=1}^{21}\left[\prod_{j=1}^{2048}\left[\begin{bmatrix}(n_{01}n_{12}\dots n_{0j})\\(n_{11}n_{12}\dots n_{1j})\\ \vdots\\ \vdots\\(n_{N1}n_{N2}\dots n_{Nj})\end{bmatrix}\right]\right]$$

Where, i is some images, and each image is converted to 2,048 features. We need to consider each of these images separately to identify the customer's sentiments.

On the other hand, Lexicon analysis espouses the automated evaluation of sentiment analysis, and is thus commonly accepted (Wang, Li, Huang, Liu & Zhang, 2017). Ekman model (t) matches the term extracted of the 2,048 features extracted for each image.

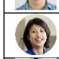
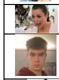
Finally, Liu Hu Method (Hu & Liu, 2004; Hu, Bose, Koh & Liu, 2012; X. Hu, Tang, Tang & Liu, 2013) is used for visual sentiment analysis polarity of image.

$$polarity(t) = \begin{cases} happy \sum_{p \in P} weight_{happy}(p) > \sum_{n \in N} weight_{sad}(n) \\ sad \sum_{p \in P} weight_{happy}(p) < \sum_{n \in N} weight_{sad}(n) \end{cases}$$

Note that in an equation, p and n are happy and sad emotions. Also, p and n are the obtained emotions using the 2,048 features for each image. Neutral polarity does not form the study's part; instead, the 2 types of polarity form the study's foundation to identify the sentiments of social media reactions via images.

Part 1 of the proposed model's outcomes are shown in Table 1, where there are 2 classifications of images, f1 = happy and f2 = sad.

Table 1: Outcomes of Part 1 for Image Classification

Image	Category
	f1 f2
	Class (f1 = happy, f2 = sad)
	f1
	f1
	f1
	f1
	f1
	f1
	f1
	f1
	f1
	f1
	f2
	f2
	f2
	f2
	f2
	f2
	f2
	f2
	f2
	f2
	f2

Part 2: Classification Algorithms for Visual Sentiment Analysis

As discussed, comparative analysis of four machine learning classification algorithms is carried out. The mathematical model for fake news detection is elaborated in the section.

Support Vector Machine

Support vector machine is a supervised classification algorithm that classifies the emotions into happy and sad, based on the hyperplane. The equation of hyperplane is given as follows.

$$H: w^{T}(x) + b = 0$$

Where, b is the intercept and bias term of the hyperplane equation. There are 2 challenges faced by the support vector machine.

- While fitting the separating line, we require the line that would be able to segregate the emotions as happy and sad in the best possible way.
- Secondly, we also need to make sure that there is the slightest errors in the classification between happy and sad emotions.

Both of the 2 points is validated by calculating the distance as follows.

$$d_H(\emptyset(x_0)) = \frac{|w^T(\emptyset(x_0)) + b|}{\|w\|_2}$$

Where, $\|w\|_2$ is the Euclidean distance.

$$\|w\|_2 = \sqrt{w_1^2 + w_2^2 + w_3^2 + \cdots w_n^2}$$

The goal of the SVM algorithm is to maximise the minimum distance of emotions from the hyperplane.

$$w^* = arg_w max[min_n d_H(\emptyset(x_n))]$$

While making the predictions on the image dataset, which were binary classified as happy and sad, if the point substitutes from happy emotions in the hyperplane equation, we get a value greater than 0 (zero). Mathematically,

$$w^T(\emptyset(x)) + b > 0$$

And predictions from the sad emotions in the hyperplane equation would give a negative value,

$$w^T(\emptyset(x)) + b < 0$$

Naïve Bayes

In Naïve Bayes, we transform the matrix into a vector, as mentioned in the equation, considering A and B as 2 events for sample space.

$$P(class|feature\ set) = \frac{P(feature\ set|class)P(class)}{P(feature\ set)}$$

The probability *P*(*class*|*feature set*) is called posteriori, gives the probability of classifying the image as happy or sad, given a set of features that can be observed as happy or sad from the Inception V3 algorithm. *P*(*class*) is called prior, as it has all the information we have, the probability of being happy or sad. *P*(*feature set*) is called evidence and is the probability of what we are observing, the set of features. *P*(*feature set*| *class*) is the likelihood of an image as happy or sad with specific set of features.

Evaluating the posteriori probability, between *P*(*feature set*| happy) or *P*(*feature set*| sad) maximizes the likelihood of image in either of the class.

$$\widehat{class} = argmax_{class} P(Class = class \mid Feature\ set = feature\ set)$$

Neural Network

The algorithm classifies extracted emotions like happy and sad (binary classifier labelled as 1 for happy emotions and 0 for sad emotions) based on features. The data is grouped into a matrix, where each column corresponds to a component and each row represents a single data point and is defined as $X, X \epsilon \mathbb{R}^{(mxn)}$. We then have a vector, $\hat{y}, \hat{y} \epsilon \mathbb{R}^m$, which contains the outputs, either 0 for sad or 1 for happy. For neural networks, we also define weight vector (to adjust the values as per the cost function) $w, w \epsilon \mathbb{R}^n$. The weighted sum is:

$$w_1 x_1 + w_2 x_2 + \cdots \ldots \ldots \ldots \ldots .. w_m x_m$$

which is given as

$$\sum_{i=1}^{m} w_i x_i = w^T x$$

Where w^T is "*transpose w*."

Logistic regression uses probabilistic logistic function, based on the equation mentioned above. It is given as follows.

$$P(w^T x) = \frac{1}{1 + e^{-w^T x}}$$

Based on the equation, if the weighted sum for a data point is nearing 1, then we can predict thedata point to have a class 1; it is 0 otherwise, which is given as follows.

$$P(\hat{y} = 1|x: w) = \frac{1}{1 + e^{-w^T..}}\ for\ class\ 1$$

For correct classification, Stochastic Gradient Descent (SGD) iterates the cost function as

$$C = \frac{1}{2}\sum(\hat{y} - y)^2$$

Initially, $y = 0$.

Random Forest

Random forest is based on a decision tree concept and evaluates entropy and gain at each level of the tree.

Entropy is the measure of the randomness of a system and is used for the training set as follows.

$$Entropy\ E(S) = -\sum p(x) * log_2(p(x))$$

Where, (x) gives the probability of getting the class (happy or sad) from the training dataset. Overall, the entropy for the algorithm is

$$Entropy\ E(S) = -p(happy) * log_2(p(happy)) + p(sad) * log_2(p(sad))$$

The gain is calculated based on the decrease in entropy after a dataset gets split on an attribute.

$$Gain(happy, Inception\ V3) = E(S) - \sum_{Inception\ V3 \epsilon D_{Inception\ V3}} \frac{|S_{happy}|}{|S|} Entropy(S_{happy})$$

$$Gain(sad, Inception\ V3) = E(S) - \sum_{Inception\ V3 \epsilon D_{Inception\ V3}} \frac{|S_{sad}|}{|S|} Entropy(S_{sad})$$

Classification Algorithm Outcome Parameters

The model built using the 21 images dataset was tested for their prediction accuracies. We evaluated 4 performance statistics for the developed model:

CA: Classification Accuracy

Classification accuracy is the fraction of predictions the model we developed got right (Alić et al., 2017; Wu et al., 2012).

$$CA = \frac{Number\ of\ correct\ predictions}{Total\ number\ of\ predictions}$$

For binary classifiers:

$$CA = \frac{TP + TN}{TP + TN + FP + FN}$$

Typically, we consider $CA > 0.5$ as acceptable.

Precision:

Precision is the estimation of TP for positives.

$$Precision = \frac{TP}{TP + FP}$$

We aspect for $Precision > 0.5$.

Recall:

Recall calculates TP among all positive instances.

$$Recall = \frac{TP}{TP + FN}$$

We look for $Recall > 0.5$.

F1:

F1 is the weighted harmonic mean for precision and recall, and creates a balance between precision and recall (Brownlee, 2014).

$$F1 = 2 * \frac{Precision * Recall}{Precision + Recall}$$

We take $F1 > 0.5$.

The proposed model's prediction accuracy for each of the 4 classification machine learning models is shown in Table 2.

Table 2: Research Outcomes for Classification Models

Models	*CA*	*Precision*	*Recall*	*F1*
SVM	0.810	0.860	0.810	0.800
Random Forest	0.619	0.627	0.619	0.617
Neural Network	0.857	0.860	0.857	0.856
Naïve Bayes	0.810	0.820	0.810	0.807

The prediction accuracy of 21 image datasets displays the model's strength, which postulates no overfitting of the model. All they did were the correct classifications, with the acceptable margins. However, the neural network model predicted most of the emotions correctly, with almost 86% accuracy.

Conclusion

Visual sentiment analysis on social network content can help one recognise consumer behaviour and provide helpful information for related data analysis. This paper introduced a novel visual sentiment analysis model using Inception V3 for image embedding and Liu Hu and Ekman's model for sentiments and emotions exploration. We tested the model using four classification machine learning models to evaluate the accuracy of the Inception V3 algorithm. Neural network classification predicted the images correctly, with 86% accuracy.

References

Alam, F., Imran, M., & Ofli, F. (2017). Image4Act: Online social media image processing for disaster response. *Proceedings of the 2017 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining 2017 – ASONAM '17*, pp. 601-604. doi:https://doi.org/10.1145/3110025.3110164

Almgren, K., Kim, M., & Lee, J. (2017). Mining social media data using topological data analysis. *2017 IEEE International Conference on Information Reuse and Integration (IRI)*, pp. 144-153. doi:https://doi.org/10.1109/IRI.2017.41

Amato, G., Bolettieri, P., Monteiro de Lira, V., Muntean, C. I., Perego, R., & Renso, C. (2017). Social media image recognition for food trend analysis. *Proceedings of the 40th International ACM SIGIR Conference on Research and Development in Information Retrieval*, pp. 1333-1336. doi:https://doi.org/10.1145/3077136.3084142

Appel, G., Grewal, L., Hadi, R., & Stephen, A. T. (2019). The future of social media in marketing. *Journal of the Academy of Marketing Science*. doi:https://doi.org/10.1007/s11747-019-00695-1

Bakhshi, S., Shamma, D. A., & Gilbert, E. (2014). Faces engage us: Photos with faces attract more likes and comments on Instagram. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pp. 965-974. doi:https://doi.org/10.1145/2556288.2557403

Basnyat, B., Anam, A., Singh, N., Gangopadhyay, A., & Roy, N. (2017). Analyzing social media texts and images to assess the impact of flash floods in cities. *2017 IEEE International Conference on Smart Computing (SMARTCOMP)*, pp. 1-6. doi:https://doi.org/10.1109/SMARTCOMP.2017.7946987

Castiglione, A., Cattaneo, G., & De Santis, A. (2011). A forensic analysis of images on online social networks. *2011 Third International Conference on Intelligent Networking and Collaborative Systems*, pp. 679-684. doi:https://doi.org/10.1109/INCoS.2011.17

Corliss, R. (2019). Photos on Facebook generate 53% more likes than the average post [New Data]. Retrieved October 23, 2019, from https://blog.hubspot.com/blog/tabid/6307/bid/33800/photos-on-facebook-generate-53-more-likes-than-the-average-post-new-data.aspx

Garimella, V. R. K., Alfayad, A., & Weber, I. (2016). Social media image analysis for public health. *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*, pp. 5543-5547. doi:https://doi.org/10.1145/2858036.2858234

Hu, M., & Liu, B. (2004). Mining and summarizing customer reviews. *Proceedings of the Tenth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pp. 168-177. doi:https://doi.org/10.1145/1014052.1014073

Hu, N., Bose, I., Koh, N. S., & Liu, L. (2012). Manipulation of online reviews: An analysis of ratings, readability, and sentiments. *Decision Support Systems*, *52*(3), 674-684.https://doi.org/10.1016/j.dss.2011.11.002

Hu, X., Tang, L., Tang, J., & Liu, H. (2013). Exploiting social relations for sentiment analysis in microblogging. *Proceedings of the Sixth ACM International Conference on Web Search and Data Mining*, pp. 537-546. doi:https://doi.org/10.1145/2433396.2433465

Islam, J., & Zhang, Y. (2016). Visual sentiment analysis for social images using transfer learning approach. *2016 IEEE International Conferences on Big Data and Cloud Computing (BDCloud), Social Computing and Networking (SocialCom), Sustainable Computing and Communications (SustainCom) (BDCloud-SocialCom-SustainCom)*, pp. 124-130. doi:https://doi.org/10.1109/BDCloud-SocialCom-SustainCom.2016.29

Ji, R., Cao, D., Zhou, Y., & Chen, F. (2016). Survey of visual sentiment prediction for social media analysis. *Frontiers of Computer Science*, *10*(4), 602-611. doi:https://doi.org/10.1007/s11704-016-5453-2

Maigrot, C., Kijak, E., & Claveau, V. (2018). Context-aware forgery localization in social-media images: a feature-based approach evaluation. *2018 25th IEEE International Conference on Image Processing (ICIP)*, pp. 545-549. doi:https://doi.org/10.1109/ICIP.2018.8451726

Mantyla, M. V., Novielli, N., Lanubile, F., Claes, M., & Kuutila, M. (2017, May). *Bootstrapping a Lexicon for Emotional Arousal in Software Engineering*, pp. 198-202. doi:https://doi.org/10.1109/MSR.2017.47

Pittman, M., & Reich, B. (2016). Social media and loneliness: Why an Instagram picture may be worth more than a thousand Twitter words. *Computers in Human Behavior*, *62*, 155-167. doi:https://doi.org/10.1016/j.chb.2016.03.084

Tsang, S.-H. (2015). Inception-v3—1st Runner Up (Image Classification) in ILSVRC 2015. Retrieved from https://medium.com/@sh.tsang/review-inception-v3-1st-runner-up-image-classification-in-ilsvrc-2015-17915421f77c

Wang, W., Li, Y., Huang, Y., Liu, H., & Zhang, T. (2017). A method for identifying the mood states of social network users based on cyber psychometrics. *Future Internet*, *9*(2), 22. doi:https://doi.org/10.3390/fi9020022

Wang, Y., & Li, B. (2015). Sentiment Analysis for Social Media Images. *2015 IEEE International Conference on Data Mining Workshop (ICDMW)*, 1584-1591. https://doi.org/10.1109/ICDMW.2015.142

Webster. (2015). 8 Surprising Twitter Statistics to Get More Engagement [Postcron]. Retrieved from Postcron—Social Media Marketing Blog and Digital Marketing Blog website: https://postcron.com/en/blog/8-surprising-twitter-statistics-get-more-engagement/