Using Machine Learning to Build an Emotion Recognizer for Images

Abstract

Emotion-aware news production, in conjunction with audience profiling capability, can enable media companies to deepen customer engagement. Superior engagements, in turn, increase reader loyalty, and create significantly more opportunities for companies to sell ad space.

Far from being a nice-to-have, the emotion recognition tool has become a critical requirement for businesses across the media, communication, and information industry. The ability to harness instant recognition of emotions from multiple channels can be a differentiator for modern businesses that need to constantly find the right audiences at the right time.

Instead of building a model for generic situations, organizations need to create one tailored to their business needs. Thus, an emotion recognizer designed for a mass media organization should be technically different from that built for a publishing house. Such a customized approach will help organizations achieve higher accuracy with greater cost savings.

In this paper, we discuss key lessons that help companies realize the desired outcomes from an emotion recognition system, based on machine learning technologies.

The Importance of Analyzing Emotions in Digital Images

In their popular paper, 'What Makes Online Content Viral?[1],', professors Jonah Berger and Katherine Milkman of the Wharton School, University of Pennsylvania examined how emotions catalyze viral effects. By analyzing a series of New York Times articles published over a three-month period, they found positive content to be more viral than negative content. The authors also highlighted the complex relationship between emotion and social transmission, attributing it partially to physiological arousal.

An opinionated, thought provoking article inducing awe or anger might typically appeal more to a left-leaning, educated, middle-aged, well-to-do, urban populace. The same article might not make any emotional connect with readers who are younger, earn less money, and come from a different educational background. Though it cannot be generalized that more anger will bring in more readership, it is certain that a strong emotional connect can significantly influence sales.

A predominant element in the equation between emotion and sales is images, which can evoke certain emotions faster than non-visual elements. Furthermore, emotions can be quickly transferred through images without much textual explanation. Accurate identification and analysis of emotions conveyed through digital images have gained increased importance for businesses striving to implement effective customer experience initiatives. Therefore, building an emotion recognition system for images is essential for an emotional context-aware business model.

In this paper, we discuss key lessons media organizations should keep in mind while building an emotion recognizer for digital images containing facial expressions.

Building a corpus

A corpus of images capturing human facial expressions provides the source of the intelligence for the automatic emotion-recognition machine. The key is to create a large library of images at the beginning of the project lifecycle by labeling each image with an emotion.

Determining what emotions would be ideal for a business is an important step to begin with. For instance, a news publisher would need to categorize emotions like angery, sadness, and happiness, etc. to be able to connect images to news articles, based on the emotion articles trigger. On the other hand, music companies might require a completely different set of psychological attributes like romanticism, relaxationing, and joyfulness, etc. for their specific business environment. The total number of emotions that may seem necessary for one business may vary for another business.

Based on the pioneering research of American psychologist Paul Ekman, there are seven categories[2] of human emotion considered universal for facial images–, including joy, sadness, anger, disgust, contempt, surprise, and fear , – it is likely two teachers will end up categorizing the same image differently. To keep subjective bias under control, it is advisable to repeat the tagging exercise through multiple teachers, and aggregate individual inputs. By identifying the emotion category that receives the highest votes, it is easier to confirm the category as the appropriate emotion label of the image.

Since labeling involves significant amount of time and effort, enterprises need to provision these in their budget and schedule to ensure the success of the entire project.

Streamlining data preprocessing

In the context of image recognition, data consists of a collection of digital images of individual faces. Before data is fed into the machine learning system, it goes through the following key processes:

Noise reduction

There might be a few blurred or tilted images that are removed from the collection that gets built for training. Images that are too dark or too bright can be corrected using the 'equalize histogram' method. The entire collection of images are checked and preprocessed cleaning non-recognizable faces.

Data exploration

It is important to visualize the emotion of the mean face. The mean face is synthetically derived combining all the images of an emotion group. The mean image is generated using the mean of pixel intensities. The mean face provides a visual affirmation of the particular emotion category. The extent of emotion resemblance is an indicator of the quality of training data.

Data augmentation

Considering that an image recognition model cannot discover inherent patterns from inadequate volume of data, there are data augmentation techniques that can be considered to increase the training data volume including image multiplication, image flipping, and affine transformation.

Feature extraction

A way to recognize emotion from a facial expression is to study the face animation parameters that correspond to the actions of facial muscles. This essentially considers the relative distances of the key facial points. To determine the key points of an image, eyes, eyebrows, mouth, and nose are detected using image processing and computer vision techniques based on Sobel edge and Harris corner detectors.

Building the intelligent recognizer

Once the key facial points are identified, a set of Euclidean distances are determined for each image from various combinations of key points. The distances are then normalized with respect to the length and breadth of the face. Consequently, each image gets represented by the set of normalized distance measures.

It is recommended to build multiple models using machine learning algorithms for Support Vector Machine (SVM), Random Forests, Multi-Layer Perceptron (MLP) and Convolutional Neural Network (CNN). Each model requires individual training and tuning to bring out the best suited model parameters for the given training set.

We suggest using the Ensemble model as the final classifier. This method identifies high performing models by assigning them higher votes. It combines all the models, and works as the classifier for prediction of emotion.

Conclusion

Capability to invoke emotion through images and understanding the emotion invoked by an image is invaluable to many companies, particularly those in the media industry. Typically, the business driver to build an emotion recognizer is improved customer experience, which ties in very closely with the digital agenda many organizations are pursuing today. Choosing an appropriate emotion recognizer solution ensures the consistent delivery of the expected business outcomes. By harnessing the power to instantly recognize emotions across multiple channels, businesses can reap the true benefits of machine learning, ultimately broadening and deepening customer engagements.

References

1. Journal of Marketing Research, What Makes online Content Viral? (2011), accessed February 2017, http://jonahberger.com/wp-content/uploads/2013/02/ViralityB.pdf

2. Paulekman, Facial Expression of Emotion (2000), accessed May 2017, https://www.paulekman.com/wp-content/uploads/2013/07/Universal-Facial-Expressions-Of-Emotion.pdf

TATA CONSULTANCY SERVICES

About The Authors

Priyankar Ghosh

Priyankar Ghosh leads the Machine Learning Group of the Content Lab, a Communication, Media and Information (CMI) research and innovation center at Tata Consultancy Services Limited (TCS), Kolkata, India. Ghosh focuses on research in AI, machine learning, data science, and natural language processing. He has over 19 years of experience in the IT industry, spanning areas such as research, consulting, and delivery management, and holds an M.Stat. degree from the Indian Statistical Institute.

Kathakali Seth

Kathakali Seth is a member of the Machine Learning Group of the Content Lab. Her research interests pertain to areas such as machine learning, data mining, data visualization, and data science. With over 19 years of experience in the IT industry, she specializes in building and implementing applications. Seth holds a B.E. degree from the Indian Institute of Engineering Science and Technology.

Anjan Dutta

Anjan Dutta is the Principal Research Engineer of the Content Lab. His research interests are in natural language processing, text analytics, and machine learning. He brings to the table over 22 years of experience in the IT industry, spanning areas such as enterprise architecture, software development, research, and consulting. He holds a B.Tech. degree from the Indian Institute of Technology, Kharagpur and Master of Engg. (Collaborative) from BITS, Pilani.

Experience certainty. IT Services
Business Solutions
Consulting

Contact

Visit the Communications, Media & Technology page on www.tcs.com

Email: global.cmi@tcs.com

Subscribe to TCS White Papers

TCS.com RSS: http://www.tcs.com/rss_feeds/Pages/feed.aspx?f=w

Feedburner: http://feeds2.feedburner.com/tcswhitepapers

About Tata Consultancy Services Ltd (TCS)

Tata Consultancy Services is an IT services, consulting and business solutions organization that delivers real results to global business, ensuring a level of certainty no other firm can match. TCS offers a consulting-led, integrated portfolio of IT and IT-enabled, infrastructure, engineering and assurance services. This is delivered through its unique Global Network Delivery Model™, recognized as the benchmark of excellence in software development. A part of the Tata Group, India's largest industrial conglomerate, TCS has a global footprint and is listed on the National Stock Exchange and Bombay Stock Exchange in India.

For more information, visit us at www.tcs.com

TCS Design Services I M I 07 I 17