A Framework for Facial Image Analytics Using Deep Learning in Social Sciences Research

Stuart J. Barnes[1( )] and Richard Rutter[1,2]

[1] CODA Research Centre, King's Business School, King's College London, London, UK
stuart.barnes@kcl.ac.uk richrutter@gmail.com
[2] School of Business, Australian College of Kuwait, Mishref, Kuwait

Abstract Recent advances in arti cial intelligence have provided exciting new tools for scienti c research that have been applied in many domains, including medicine, engineering, nance, the physical sciences, and online consumer analytics. Although text analytics is now quite a mature approach, image analytics is still developing as research approach and its use is not widespread in the social sciences, particularly owing to the specialized knowledge required in visualization and machine learning. In this paper, we demonstrate how facial image analytics research can be expedited in the social sciences by introducing a simple research framework for integrating facial image classi cation data into studies. We explain the steps of the framework, including data collection, processing, collation, and analysis. This is accompanied by an example piece of research examining the in uence of hosts' facial expressions on review scores in Airbnb.

Keywords: Faces Classi cation Data fusion Deep learning Social sciences

1 Introduction

Big data analytics and machine learning provide signi cant new research opportunities for social scientists focused on the digital economy, offering large volumes of rapidly accumulating data, such as text, audio, image and video, from varied sources, e.g. mobile devices, sensors (including internet of things), open or public data, online social networks, and organizational information systems (Baesens et al. 2016; Goes 2014). The majority of big data is unstructured (with estimates ranging from 80 to 95 or more; see Gandomi and Haider 2015), including much that is audio, images and video. Although big data research with text analytics has reached maturity and is being widely applied in the social sciences (e.g. see Villarroel Ordenes et al.'s 2017 paper on the examination of customer sentiment), research utilizing images seldom used, being con ned to a niche of experts with knowledge of computer visualization and machine learning. A few rare examples include Zhang and Luo (2018) and Wang et al. (2018). This paper aims to provide social scientists with a simple framework to enable advanced facial analytics in new academic investigations.

Academic social science research utilizing machine learning is a very different proposition to typical business applications of the methods, which are typically

R. Jallouli et al. (Eds.): ICDEc 2019, LNBIP 358, pp. 315 320, 2019.
https://doi.org/10.1007/978-3-030-30874-2_25

employed to solve practical problems without mentioning theory. Such problems include predicting customer demand to enhance the ef ciency of supply chains, predicting customer propensity to churn to improve approaches to customer retention, examining preferences to make customer product recommendations and increase sales, optimizing pricing to grow pro ts, fraud detection to lower risk, and personalization to improve customer loyalty. In the next section, we introduce steps in the research framework, illustrating its application to examine the impact of facial expressions of emotion on online reviews for Airbnb, using a freely available online data source. In the nal section we draw conclusions on the opportunities and limitations for the future application of the research framework in social science research.

2 Research Framework

In this section, we examine the recommended research framework for facial image analysis in social sciences by outlining six key steps in the research process (illustrated in Fig. 1). The research framework generally ts with other generalized processes for big data analytics (e.g. Gandomi and Haider 2015). We examine each step, in turn, focusing on an example application.

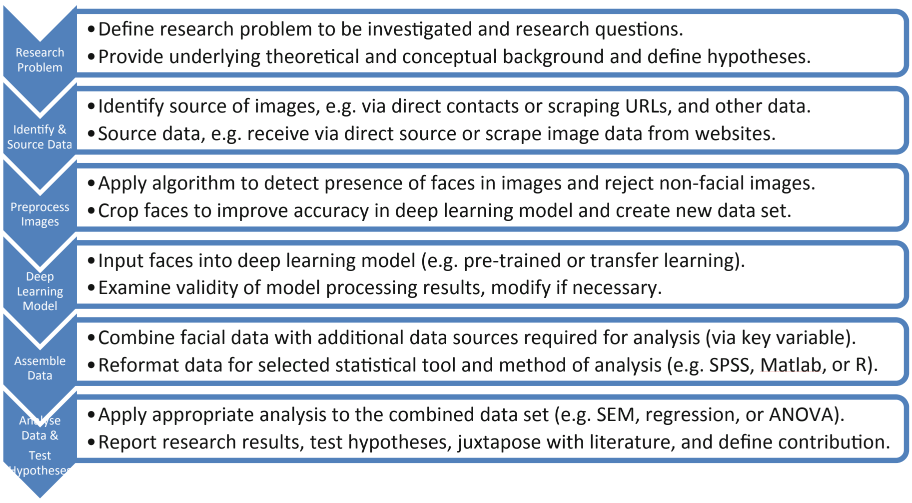

Fig 1 Steps in the recommended process for facial image analysis

2 1 Research Problem

A key step in developing facial analytics research that makes a clear contribution to knowledge is outlining the research problem, including its motivation (both academic and practical), and articulating the research question to be answered. This should be theoretically and conceptually embedded through background literature, clear definition and justification of research hypotheses to be empirically tested, and often the development of a summary research model.

In our example application, we seek to answer the research question, "Are customers' online reviews influenced by facial expressions of emotion of accommodation hosts?" When writing an online review after experiencing accommodation, customers are typically presented with an image of the host. According to emotional contagion theory (Hatfield et al. 1993), we would expect facial expressions of emotion to provide a non-verbal form of communication that could indirectly and automatically (Hatfield et al. 1993), influence the customer in their evaluation of the accommodation. We therefore hypothesize that:

H1. Customers' overall review evaluations will vary according to the emotion of the accommodation host.

2.2 Identify and Source Data

Suitable sources of data to provide the empirical basis for testing the research hypotheses should be identified – high quality data that will enable clear extraction of facial features using available algorithms. In some cases, this data may be provided by open access or via public sources, or requested from researchers or organizations, in others, the data may be commercially available or potentially scraped from websites. If the images are not easily available then it may be possible to identify images using scraping software from R, Python, MATLAB and others. Once the image URLs are found, they can be downloaded using a bulk images downloader in most browsers.

In our example piece of research, we identified Airbnb as a potential example that could be used to test our hypothesis. The website includes accommodation and host information and images, as well as reviews and ratings. From a pragmatic point of view, all of the data is provided for many cities in a processed format with additional tools on the website, insideairbnb.com. The data is available under a Creative Commons CC0 1.0 Universal (CC0 1.0) Public Domain Dedication license. We decided to focus on the city of Amsterdam (the first city listed on InsideAirbnb) and used the provided URLs to download host images for approximately half of the available properties as of 5th October 2018 (n = 8056).

2.3 Preprocess Images

All of the data used in machine learning typically needs to be examined and, if necessary, preprocessed, including numbers, text, images, and so on. Traditional numeric data can be examined via visualization and statistical analysis, and may need treatment for missing values, outliers, standardization/normalization, or reclassification. It may also be necessary to sample the data. Text analysis often requires preprocessing to improve computation, including changing to lower case, tokenizing, and removing punctuation and infrequent words. In the case of images, data that is not suitable for analysis needs to be screened out, and the images need to be formatting in a way that facilitates analysis, such as cropping the face. For faces, this procedure was facilitated by applying the Viola-Jones (Viola and Jones 2001) algorithm, which systematically examines parts of the image with a cascade of binary classifiers for the face and sub-features and rejects if facial features are not found. By applying a classification model that is based on upright and forward-facing facial features, such as Haar, with a classification and regression tree (CART) analysis, we are able to screen out images that do not contain faces or that are obscured, e.g. via lighting or objects. Other types of analysis such as local binary patterns are more forgiving and could potentially detect more detail at the expense of false positives (Ojala et al. 2002). Python and MATLAB both include implementations to allow automated image processing using the Viola-Jones algorithm and cropping. A total of 2079 images were screened out as not containing faces in this process, resulting in 5977 images. After removing listings without review ratings, this fell to 5481 images.

2.4 Deep Learning Model

A plethora of methods are available for facial image analytics, including some of the most successful methods that are appearance-based and involve statistical analysis and machine learning to find key characteristics of facial images, such as Eigenface, support vector machines, principal components analysis with Fisher's discriminant, and neural networks (Masi et al. 2018). For example, Levi and Hassner (2015a, b) develop convolutional neural networks (CNNs) for accurately classifying facial images according to gender, age and emotional expression. Although it is possible to train a CNN to identify facial emotions from an image data set entirely from scratch, numerous pretrained models are available for facial emotion classification, including network structures and pretrained weights. Many advanced CNNs for image analytics are freely available in statistical software packages (e.g. GoogLeNet, ResNet-50, VGG19, and Inception-ResNet-v2) or via internet sources such as Github or the Caffe Model Zoo. CNNs are the most advanced, accurate methods for face recognition (Masi et al. 2018).

If the images for a piece of research are similar to those used for training the original network model being used, and the classification is identical, it can often be employed successfully 'straight out of the box' as a fixed feature extractor. If the images and/or classes are different, however, then the model will need to be retrained and/or retuned using transfer learning. Results must always be checked for accuracy and it may also be necessary to test various CNNs using training and validation to identify the most accurate classifier for a data set.

As an example, we focus on one of the CNNs that is freely available via the Caffe Model Zoo for facial emotion detection (any other face detection model could be substituted). One of the most accurate networks in Levi and Hassner (2015b) was found to be the CNN with local binary patterns at a radius of 5 (LBP5) coded for cyclic codes when using the VGG_S network. The pretrained CNN consisted of 24 layers for classifying the seven facial emotions examined in this study (neutral, happy, sad, fear, disgust, anger and surprised). Color (RGB) images needed to be resized to 224 x 224 pixels for the input layer. Applying the deep classification model to our data set resulted in a large number of faces classified as neutral (4077, 74.4%), with a good number of happy faces (1103, 20.1%) and sad faces (225, 4.1%). However, the other emotional classi cations had very small incidences: fear (42, 0.8), disgust (29, 0.5), and anger (5, 0.1). No faces were classi ed as surprised.

2 5 Assemble Data

Once the deep learning data has been created it needs to be combined with other research data before the analysis can begin. If the traditional data set includes the URL or name of the image, then the name of the image can be used as a key for joining the facial emotion data set with the traditional data, e.g. host pro le, review text, and review scores, in statistical packages such as SPSS, Matlab or R. In the case of our example, the Airbnb data from InsideAirbnb include the host image URL from which the image name was extracted to enable merging the data sets. Further to joining the data sets, the data may also need a further round of reformatting, depending on the intended type of statistical analysis. In our example, the variable with the strings for the dominant identi ed emotion in images needed to be converted to binary variables for the regression analysis.

2 6 Analyse Data and Test Hypotheses

Many forms of analysis can potentially be used to test the research hypotheses, depending on the nature of the hypotheses and the data being used. For example, testing for difference might employ ANOVA, whilst examining the signi cance of statistical relationships could use a suitable form of regression or structural equation modelling.

To analyze the assembled data set and test the research hypothesis regarding the impact of facial emotions on review scores, we used stepwise linear regression (probability-of-F-to-enter <=0.05 and probability-of-F-to-remove >=0.10) with the average review score as the dependent variable and each of the facial emotion binary variables as independent variables. Only a single variable was found be signi cant in the model (constant, = 95.454, p < .001): the neutral emotion had a coef cient of 0.396 and a p-value of 0.035. This appears to suggest that host images without emotion (neutral) tended to fare less well in reviews than those with some form of emotion (positive or negative, although the issue of valence is not clear). Thus, we can state that the data supports H1. However, this nding represents a very small contribution to review scores (R = 0.028).

3 Conclusions

The framework for facial image analytics using deep learning discussed in this paper provides a simpli ed, workable process for integrating advanced image analytics into social sciences research. The research approach is not without its limitations. Deep learning can be time-consuming and computationally intensive. In addition, classi cation accuracy can vary signi cantly according to data set and network employed, suggesting that it is often prudent to use transfer learning for new data sets and to compare the performance of different networks, including new ones that may emerge.

Further, the host images were not necessarily uploaded from the beginning of the posting. To increase accuracy, the upload date could be used to lter the review date.

This research was preliminary by design and conducted on subset of data. The full study will include a more comprehensive dataset. The level of accuracy of recognition is reliant on the level of accuracy within the trained networks used. A larger scale study could therefore be used to train a new or existing network and increase accuracy. Further, this research was conducted on images only and did not take into account any lexical information. A further study is planned to analyse textual data to develop greater understanding of the relationship between facial images and text (for example, the sentiment of reviews). Further, more exible approaches to image analytics might also provide better task accuracy, e.g. MaskRCNN could be used to conduct combined face detection and analysis. We hope that this research encourages other researchers to being exploring the potential for these new, advanced techniques in digital economy research in the social sciences.

References

Baesens, B., Bapna, R., Marsden, J.R., Vanthienen, J., Zhao, J.L.: Editorial: transformational issues of big data and analytics in networked business. MIS Q. **40**(4), 807 818 (2016)

Gandomi, A., Haider, M.: Beyond the hype: big data concepts, methods, and analytics. Int. J. Inf. Manag. **35**(2), 137 144 (2015)

Goes, P.: Editor's comments: big data and IS research. MIS Q. **38**(3), 3 8 (2014)

Hat eld, E., Cacioppo, J.T., Rapson, R.L.: Emotional contagion. Curr. Dir. Psychol. Sci. **2**(3), 96 99 (1993)

Levi, G., Hassner, T.: Age and gender classi cation using convolutional neural networks. In: Proceedings of the IEEE Workshop on Analysis and Modeling of Faces and Gestures (AMFG), Boston (2015a)

Levi, G., Hassner, T.: Emotion recognition in the wild via convolutional neural networks and mapped binary patterns. In: Proceedings of the ACM International Conference on Multimodal Interaction (ICMI), Seattle (2015b)

Masi, I., Wu, Y., Hassner, T., Natarajan, P.: Deep face recognition: a survey. In: Proceedings of the Conference on Graphics, Patterns and Images (SIBGRAPI), Parana, Brazil, October (2018)

Ojala, T., Pietikäinen, M., Mäenpää, T.: Multiresolution gray-scale and rotation invariant texture classi cation with local binary patterns. IEEE Trans. Pattern Anal. Mach. Intell. **24**(7), 971 987 (2002)

Villarroel Ordenes, F., Ludwig, S., De Ruyter, K., Grewal, D., Wetzels, M.: Unveiling what is written in the stars: analyzing explicit, implicit, and discourse patterns of sentiment in social media. J. Consum. Res. **43**, 875 894 (2017)

Viola, P., Jones, M.J.: Rapid object detection using a boosted cascade of simple features. In: Proceedings of the 2001 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, vol. 1, pp. 511 518 (2002)

Wang, Y., Guo, Y., Song, J.: Using image-based and text-based information for sales prediction: a deep neural network model. In: Proceedings of the American Conference on Information Systems, New Orleans, August (2018)

Zhang, M., Luo, L.: Can user generated content predict restaurant survival? Deep learning of yelp photos and reviews. Available at Social Science Research Network (2018). https://ssrn.com/abstract=3108288. Accessed 23 Nov 2018