Leveraging Cognitive Computing for Gender and Emotion Detection

Andrea Corriga[1], Simone Cusimano[1], Francesca M. Malloci[1], Lodovica Marchesi[1] and Diego Reforgiato Recupero[1]

Department of Mathematics and Computer Science,
University of Cagliari, Via Ospedale 72, 09124, Cagliari
me@andreacorriga.com, s.cusimano@studenti.unica.it,
francesca.mll@live.it, lodo.marchesi@gmail.com,
diego.reforgiato@unica.it

Abstract. In this paper we present a tool that performs two tasks: given an input image, first, (i) it detects whether the image corresponds to a male or female person and then (ii), it further recognizes which emotion the face expression of the detected person is conveying. We mapped the two tasks as multi-label classifications. The first one aims at identifying if the input image contains one of the following four categories: one male person, one female person, a group of both male and female persons or if the image does not contain any person. The second task is triggered whether the image has been recognized to belong to one of the first two classes and aims at detecting whether that image is conveying one among six different emotions: sadness, anger, surprise, happiness, disgust, fear. For both the problems, Microsoft Cognitive Services have been leveraged to extract tags from the input image. Tags are text elements that have been adopted to form the vectorial space model, using the bag of words model, that has been fed to the machine learning classifiers for the prediction tasks. For both tasks, we manually annotated 3000 images, which have been extracted from students who agreed using our system and providing their Facebook pictures for our analysis. Our evaluation uses Naive Bayes and Random Forest classifiers and with a 10-fold cross-validation reached satisfactory accuracies both for the two tasks and for the combination of them. Finally, our system works online and has been integrated with social media. In that way, any visitors logged in to Facebook through its APIs, is allowed to quickly classify any of their photos.

Keywords: Emotion Detection; Social Media; Cognitive Services; Image Classification; Machine Learning

1 Introduction

Today, around 2.45 billion people are active on different social media platforms where Facebook represents the one with the highest number of them (2 billions). Social media has become not only a key part of the modern lifestyle but also a useful marketing channel for business of all sizes. Users uploads statuses, posts, images, and videos and, therefore, this contributes to the increase of available data in Internet [17]. Most of the time, this data is made accessible only by friends or friends of friends. What not so many people know is that all the information related to a certain user might be accessible by web applications that allow associating users' Facebook account to log in within their restricted area [4]. In fact, when visiting certain websites (e.g. news), there is often the possibility to associate user's Facebook account through a dedicated pop-up, avoiding the long times needed by the registration process. It usually takes one click to perform this operation as Facebook is usually left open in users browsers. This is allowed by Facebook APIs and a validation process of the application that is being developed and, potentially, gives to the website creators access to all the information of the users. This information cannot be published as it is for privacy reasons, but it can be processed and used to tune algorithms and systems. Processing all this amount of 'big data' might be expensive but today, thanks to the development of cognitive computing tools, the elaboration (of both text and images) can be fast enough. Big data offers plenty of opportunities to unlock novel insights from the huge amount of data that is available today. Although more data (e.g. text, images, videos, sensor data) is available than ever, only a small portion of it is being analyzed and used.

Understanding emotions [2] from posts and photos is a direction that many social networks are heading: emotional disclosure can foster interpersonal connectedness and individuals are motivated to express their emotions to maintain their relatedness to others. Moreover, analyzing pictures from social media and photo-sharing websites such as Flickr, Twitter, Tumblr, Facebook can give insights into the general sentiment of people about a given event, person, organization, etc. Recent works have already investigated the mechanisms of how social network structure influences the need for emotional expression [18]. On the other hand, if we can understand the emotion an image conveys we would be able to predict emotional tags on them. There are several works that aimed at using image processing techniques for such purpose[23, 25] but not so many that solve the problem by using text description extracted from the images.

Cognitive computing systems [14, 15] are emerging and represent the third era of computing. They have been used to improve several tasks which range from sentiment analysis [24, 22], to multi-class classification of e-learning videos [5], to classification of complaints in the insurance industry [13]. Cognitive computing systems rely on deep learning algorithms and neural networks to elaborate information by learning from a training set of data. They are perfectly tailored to integrate and analyze the huge amount of data that is being released today and available. Two very well known cognitive computing systems are IBM Watson[1] and Microsoft Cognitive Services[2]. They have been employed in several domains especially within life sciences research [3].

[1] https://www.ibm.com/watson/

[2] https://azure.microsoft.com/en-us/services/cognitive-services/

In this paper we present two multi-classification problems. The first one aims at detecting whether a given image contains one male person, or one female person, or two or more people of different sex or no person. If an image belongs to one of the first two classes (is a man or a woman) one more task is performed that aims at classifying the face expression of the input image according to six possible emotions: sadness, anger, surprise, happiness, disgust, fear. Microsoft Cognitive Services have been leveraged to extract textual tags from training and test set images and create a vectorial space using bag of word model (we employed term frequency and TF-IDF) that has been fed to different classification algorithms. Our system has been validated on 3000 images manually annotated and extracted from students Facebook profiles who agreed to have their Facebook pictures used for our analysis. A 10-fold cross-validation method has been performed. We obtained an accuracy of 65% for the first classification task, an accuracy of 60% for the second task and an accuracy of 56% for the combined task. They system has also been developed to work online and allows any user to associate his/her Facebook account for logging in and classify any of the user's photo through the two tasks. The remainder of this paper is organized as it follows: Section 2 includes related work. A brief description of the Microsoft Cognitive Services is given in Section 3. Section 4 describes the used dataset and how we turned the input images into text representation. Section 5 includes the adopted methodology and the performance evaluation we carried out. Section 6 gives some technical details of our developed system. Finally, Section 7 draws conclusions and directions where we are headed.

2 Related Work

Gender recognition is a domain that has attracted interest in both fundamental and applied research. It has mostly been targeted using computer vision algorithms and its resolution is still challenging. The difficulties emerge from the different positions of a face whose capture depends on the image acquisition process, and the intrinsic differences between people's faces [19]. We turned this problem in a text classification and, to do that, we employed Cognitive Computing Systems to detect textual descriptions of the input images. The other problem we address in this paper is the Emotion detection. This problem can be tackled with computer vision techniques if we want to extract facial expression from a given image [16] or Natural Language Processing and Semantic Web techniques if we are in presence of text and want to detect or extract emotions from it [24]. In literature several works have started using Cognitive Computing Systems to extract textual (syntactic and semantic) elements that have been used to generate the vectorial space using augmentation and/or replacement techniques [5, 7, 6, 11, 12]. Results have shown to improve baselines not using such features. The opportunity that such systems offer has been therefore exploited by several approaches, also within the Sentiment Analysis domain [10, 9], that improved their accuracy and raised the competitiveness of known Sentiment Analysis challenges [8, 20, 21].

3 Microsoft Cognitive Services

We employed the Computer Vision APIs (v1.0)[3]. They provide state of the art algorithms to process images. They can be used to determine if an image contains mature content, or estimating dominant and accent colors and so on. When performing a request, the input consists of an image URL. Within the request, there is an optional parameter used to specify which features to return (image categories are returned by default). The response will be returned in JSON. Metadata we collected are: *categories, tags, description, faces, image type, color,* and *adult*. The subscription we used is free and, as such, we were able to perform only 20 requests per minute. For this reason, the online application might take several minutes to retrieve the metadata of all the images of the user. On the other hand, the Emotion APIs take a facial expression in a given input image, and returns the confidence across a set of emotions. The detected emotions are anger, contempt, disgust, fear, happiness, neutral, sadness and surprise.

4 Dataset

Our dataset set is represented by 3000 images extracted from Facebook accounts of students at University of Cagliari who agreed participating to our analysis. A subset of 1000 images contained pictures of single persons and, therefore, were used for the evaluation of the second task. These images represented face expressions having six emotions: sadness, anger, surprise, happiness, disgust, fear. The first task has been evaluated on the entire collection of images. Tables 1 and 2 shows some statistics of the collected images dataset. Moreover, images have been first sent to Microsoft Cognitive Services to retrieve the tags and manually classified according to the two classifications: one label out of four classes for the gender recognition task and one more label out of six for the emotion detection class. For the emotion detection class we followed the same procedure of the generation of the Microsoft FER dataset [1]. During the annotation, all the tags indicating the gender of a person (e.g. lady, man, woman, boy, girl) have been removed. Finally, each image has been replaced by its corresponding textual representation using tags returned by Microsoft Cognitive Services.

Table 1. Statistics for the gender detection dataset.

Images of single male persons	449
Images of single female persons	551
Images of more than one person	976
No persons	1024

[3] `https://bit.ly/2KLkskV`

Table 2. Statistics for the emotion detection dataset.

sadness	145
anger	181
surprise	166
happiness	176
disgust	180
fear	152

5 Evaluation

As mentioned within Section 4, we leveraged the textual tags extracted from Microsoft Cognitive Services to create our vectorial space model using bag of word and TF and TF-IDF models. Tags are text elements returned for each image based on more than 2000 recognizable objects, living beings, scenery, and actions. Tags are not organized as a taxonomy and no inheritance hierarchies exist. Naive Bayes and Random Forest classifiers have been employed for the obtained accuracy shown in Table 3 using 10-cross validation technique.

Table 3. Accuracies for TF and TF-IDF for the gender recognition task.

Classifier	TF	TF-IDF
Naive Bayes	65%	63%
Random Forest	60%	58%

For the second task, again, we generated the vectorial space similarly as performed above for the first task. We tested the same classifiers and in Table 4 results of the obtained accuracy are shown using 10-cross validation technique.

Table 4. Accuracies for TF and TF-IDF for the emotion detection task.

Classifier	TF	TF-IDF
Naive Bayes	62%	59%
Random Forest	58%	55%

Finally, Table 5 shows the accuracy for the combination of the two tasks: given an image, it detects if it represents a single person and, in such a case, classifies which emotion it is conveying. For simplicity, the combined task has been represented as a multi-class classification problem with 8 categories: face expression of a single person conveying each of the six emotions, more than a single person, no persons and it has been tested on the entire collection of 3000 annotated images using 10-cross validation technique.

Table 5. Accuracies for TF and TF-IDF for the combination of gender recognition and emotion detection tasks.

Classifier	TF	TF-IDF
Naive Bayes	56%	55%
Random Forest	55%	54%

The accuracy has been computed according to the following equation;

$$accuracy = \frac{tp + tn}{tp + tn + fp + fn}$$

where tp, tn, fp and fn are, respectively, true positives, true negatives, false positives and false negatives.

Table 6 shows the confusion table for the first class (male) of the gender recognition task with the quantities above (true positive, etc.) properly indicated for the computation of the accuracy. Three more accuracies values have been computed using three more similar confusion tables for the remaining three classes. The overall accuracy has been calculated averaging the four accuracies thus computed. A similar process has been used to calculate the accuracy for the emotion detection task and the combined task.

		Actual class	
		Male person.	*Non Male person*
Predicted class	*Male person*	True Positives	False Positives
	Non Male person	False Negatives	True Negatives

Table 6. Confusion table for the gender recognition task.

6 System Workflow

In this paragraph we will describe how our system works online. It is a web application composed by two main modules: the first is a web application developed using CodeIgniter Web Framework[4] which allows building robust solutions; the second module is a Python HTTP Apache Server.

We used the Facebook Graph API[5] by obtaining first the API KEY where we had to specify the needed security scopes for our application. We explicitly mentioned *user_photos* to gain the URLs of the users' public photos. During this step we had to submit a short description of our project for the Facebook team to validate. Once validated, we were able to extract the list of the public images of the user navigating our web application. Figure 1 shows the architecture of the first task[6]. A certain user visits our web application and he is redirected to

[4] `https://codeigniter.com/`
[5] `https://developers.facebook.com/docs/graph-api/`
[6] `https://bit.ly/2IrDzC1`

Facebook.com to start the authentication flow. After a successful log in, the user is redirected back to our server. The public images of the user are retrieved using a valid authentication token. The photos are sent to the Cognitive Services to fetch the related metadata. The images are shown to the user. Each photo can be clicked and a pop up appears with the related metadata and with one of the four categories related to the first task we predict on the fly with our trained classifier. In the background, a new JSON file for that user is created containing his/her images with the associated metadata and the predicted class and stored on HDFS through its REST APIs.

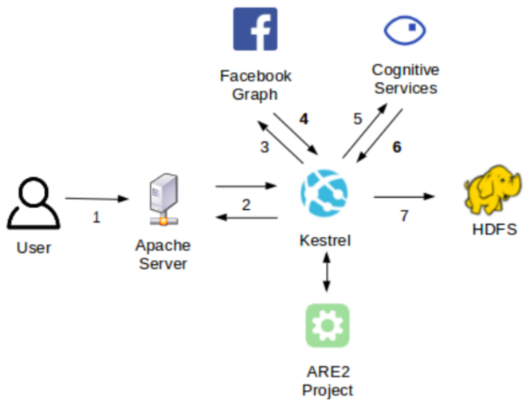

Fig. 1. Architecture of the first task.

Figure 2 shows the architecture of the second task[7] which can be run in standalone mode or once the first task has recognized the input image as a single person. In the first case, the user can decide whether to classify the image by logging in with his/her user profile through the Facebook entry or by uploading the image through the upload button. After one or more images have been chosen, a task in background starts to send the image to the Computer Vision Services and Emotion Services to fetch the related metadata. The description tags returned from Computer Vision are sent to our classification to predict the corresponding emotion. Once the computation has been completed, the user can see for that image the results of our classification related to the emotion conveyed by the face expression extracted from the input image.

7 Conclusions

In this paper we presented an approach that performs two multi-class classification tasks: detecting gender of a given image and detecting emotions of face

[7] `https://bit.ly/2jQpOzb`

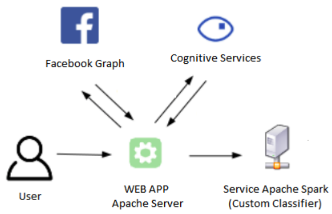

Fig. 2. Architecture of the second task.

expression for images detected within the first task. The system has been developed and works on-line and integrated with Facebook so that visitors can log in associating their Facebook account and deciding to perform the classification task on their images. For each selected image, Microsoft Cognitive Services are called and image tags collected in order to create a vectorial space representation of the input image. Our evaluation on 3000 images for the first task and 1000 images for the second task using two different classifiers with 10-cross validation generated accuracy respectively of 65% using TF and 63% using TF-IDF for the first task, 59% using TF and 55% using TF-IDF for the second task and, 56% (TF) and 55% (TF-IDF) for the combined task. As future work we are already collecting more images using crowd sourcing tools and will apply deep learning models on machines with fast NVIDIA GPUs (e.g. TitanX).

Acknowledgments

The authors gratefully acknowledge Sardinia Regional Government for the financial support (Convenzione triennale tra la Fondazione di Sardegna e gli Atenei Sardi Regione Sardegna - L.R. 7/2007 annualità 2016 - DGR 28/21 del 17.05.201, CUP: F72F16003030002).

References

1. Emad Barsoum, Cha Zhang, Cristian Canton-Ferrer, and Zhengyou Zhang. Training deep networks for facial expression recognition with crowd-sourced label distribution. *CoRR*, abs/1608.01041, 2016.
2. Davide Buscaldi and Belém Priego Sanchez. LIPN-UAM at emoint-2017: Combination of lexicon-based features and sentence-level vector representations for emotion intensity determination. In *Proceedings of the 8th Workshop on Computational Approaches to Subjectivity, Sentiment and Social Media Analysis, WASSA@EMNLP 2017, Copenhagen, Denmark, September 8, 2017*, pages 255–258, 2017.
3. Ying Chen, JD Elenee Argentinis, and Griff Weber. Ibm watson: How cognitive computing can be applied to big data challenges in life sciences research. *Clinical Therapeutics*, 38(4):688 – 701, 2016.

4. Gianpiero Costantino, Fabio Martinelli, and Daniele Sgandurra. Are photos on social networks really private? In *2013 International Conference on Collaboration Technologies and Systems, CTS 2013, San Diego, CA, USA, May 20-24, 2013*, pages 162–165, 2013.
5. D. Dessì, G. Fenu, M. Marras, and D. Reforgiato Recupero. Leveraging cognitive computing for multi-class classification of e-learning videos. In E. Blomqvist, K. Hose, H. Paulheim, A. Ławrynowicz, F. Ciravegna, and O. Hartig, editors, *The Semantic Web: ESWC 2017 Satellite Events*, pages 21–25, Cham, 2017. Springer International Publishing.
6. D. Dessì, G. Fenu, M. Marras, and D. Reforgiato Recupero. Bridging learning analytics and cognitive computing for big data classification in micro-learning video collections. *Computers in Human Behavior*, 2018. Article in Press.
7. D. Dessì, D. Reforgiato Recupero, G. Fenu, and S. Consoli. Exploiting cognitive computing and frame semantic features for biomedical document clustering. In *Proc. of SeWeBMeDA@ESWC 2017, Portoroz, Slovenia, May 28, 2017.*, pages 20–34, 2017.
8. M. Dragoni and D. Reforgiato Recupero. Challenge on fine-grained sentiment analysis within eswc2016. *Communications in Computer and Information Science*, 641:79–94, 2016.
9. Mauro Dragoni and Giulio Petrucci. A neural word embeddings approach for multi-domain sentiment analysis. *IEEE Trans. Affective Computing 8(4)*, pages 457–470, 2017.
10. Mauro Dragoni and Giulio Petrucci. A fuzzy-based strategy for multi-domain sentiment analysis. *International Journal of Approximate Reasoning*, 93:59–73, 2018.
11. Amna Dridi, Mattia Atzeni, and Diego Reforgiato Recupero. Finenews: fine-grained semantic sentiment analysis on financial microblogs and news. *International Journal of Machine Learning and Cybernetics*, 2018.
12. Amna Dridi and Diego Reforgiato Recupero. Leveraging semantics for sentiment polarity detection in social media. *International Journal of Machine Learning and Cybernetics*, 2017.
13. J Forster and B Entrup. A cognitive computing approach for classification of complaints in the insurance industry. *IOP Conference Series: Materials Science and Engineering*, 261(1):012016, 2017.
14. J. O. Gutierrez-Garcia and E. López-Neri. Cognitive computing: A brief survey and open research challenges. In *2015 3rd International Conference on Applied Computing and Information Technology/2nd International Conference on Computational Science and Intelligence*, pages 328–333, 2015.
15. John E. Kelly and Steve Hamm. *Smart Machines: IBM's Watson and the Era of Cognitive Computing*. Columbia University Press, New York, NY, USA, 2013.
16. Jyoti Kumari, R. Rajesh, and K. M. Pooja. Facial expression recognition: A survey. *Procedia Computer Science*, 58:486–491, 2015.
17. Savita Kumari. Impact of big data and social media on society. *Global Journal For Research Analysis*, 5, March 2016.
18. Han Lin, William Tov, and Lin Qiu. Emotional disclosure on social networking sites. *Comput. Hum. Behav.*, 41(C):342–350, December 2014.
19. Choon-Boon Ng, Yong-Haur Tay, and Bok-Min Goi. A review of facial gender recognition. *Pattern Analysis and Applications*, 18(4):739–755, Nov 2015.
20. D.R. Recupero, M. Dragoni, and V. Presutti. Eswc 15 challenge on concept-level sentiment analysis. *Communications in Computer and Information Science*, 548:211–222, 2015.

21. D. Reforgiato Recupero, E. Cambria, and E. Di Rosa. Semantic sentiment analysis challenge at eswc2017. *Communications in Computer and Information Science*, 769:109–123, 2017.
22. Diego Reforgiato Recupero, Valentina Presutti, Sergio Consoli, Aldo Gangemi, and Andrea Giovanni Nuzzolese. Sentilo: Frame-based sentiment analysis. *Cognitive Computation*, 7(2):211–225, 2015.
23. Can Xu, Suleyman Cetintas, Kuang-Chih Lee, and Li-Jia Li. Visual sentiment prediction with deep convolutional neural networks. *CoRR*, abs/1411.5731, 2014.
24. Ali Yadollahi, Ameneh Gholipour Shahraki, and Osmar R. Zaiane. Current state of text sentiment analysis from opinion to emotion mining. *ACM Comput. Surv.*, 50(2):25:1–25:33, May 2017.
25. Quanzeng You, Jiebo Luo, Hailin Jin, and Jianchao Yang. Robust image sentiment analysis using progressively trained and domain transferred deep networks. *CoRR*, abs/1509.06041, 2015.