Image Sentiment Analysis: Experimental Evaluation of Several Deep Learning Architectures

António P. Gaspar
antonio.pedro.gaspar@ubi.pt
Luís A. Alexandre
luis.alexandre@ubi.pt

Universidade da Beira Interior
Instituto de Telecomunicações
Rua Marquês d'Ávila e Bolama
6201-001, Covilhã, Portugal

Abstract

Image sentiment analysis is an important topic nowadays. It is possible to use it to classify an image at sentiment level, as negative, neutral or positive. However, to classify an image at this level is a hard challenge because its semantic meaning can represent many scenarios. In this paper, we present an analysis of several image classification methods that we evaluate to improve the state of the art in a large tweet data set.

1 Introduction

Sentiment analysis is a very studied subject. Currently, the data of social media networks is growing at each second, which makes them a good place to collect images. These can be analysed and classified for different purposes, such as sentiment analysis. There are methods capable to handle this job. However, the majority of these methods has a high text dependency to realise the sentiment classification. A widely used expression is "A picture is worth a thousand words", that means, that an image can transmit a large message. Nonetheless, the message is not always clear, and this situation can bring different interpretations, especially when the interpreters have different cultures. With this work, we propose an improvement of the method proposed in [4], to do sentiment analysis using only the images present in tweets. Section 2 presents related work, including the method from [4], that we partially improve in this paper. Section 3 describes our work. Section 4 contains the experiments and the final section has our conclusions.

2 Related Work

2.1 Sentiments

According to the psychology area, sentiments are different from emotions. This fact is described by the author of the paper [5]. Sentiments are the result of subjective experiences that were lived from an emotion. Emotions are the triggers for actions that can be positive or negative and are the base of sentiments. These can construct the history of the all feelings that are processed and memorised. This fact is important in sentiment analysis because through it is possible to reduce the subjectivity according to the culture where the analysed data belongs to. However, often the data cannot be organised by the culture. It is the case of the data collected from social media networks. For this reason, artificial intelligence may help to find the best features for classification. Next, we present some of the techniques related with the present theme.

2.2 Sentiments and Artificial Intelligence

Nowadays text, images, videos and all multimedia content can be processed and analysed. To analyse the data, most models represent information using sets of features which in turn represent the classes of the target objects. This process can be done through many different approaches, but currently, deep neural networks, such as Convolutional Neural Networks (CNNs), have been producing very good results when applied to image data.

There have been many proposals of methods for Image Sentiment Analysis. The authors of [7] studied the sentiment analysis process. They propose a method that is capable of classifying images at sentiment polarity level. The dataset they use is composed of 3 million tweets, which include text and images, and was constructed by them collecting the information on Twitter. For the classification, they propose a method that leverages the text classification and correlates it with the image. They conclude that text associated with image is often noisy and is weakly correlated with the image content, but it is possible to classify its sentiment using a model that is trained with the images classified with text labels.

In another work described in [2], the authors explore four different architectures of convolutional neural networks to do sentiment analysis in visual media. This work was based on a labelled set that has the main categories of the description of the scene. With their results, the authors compose their own dataset and train a model that improves the results. With this knowledge, following section present the method developed in [4].

2.3 Image Content Analysis

In [4] an image content analysis method was developed. This is a complex subject because an image might contain many objects. This work tries to identify automatically the class of the object that an image can represent. To do this a pre-trained model with the ImageNet [3] was used to classify the data into its class through the ImageNet classifications (1000 possible object classes). All images on the ImageNet are quality-controlled and human-annotated. An InceptionResNetV2 trained model was used, which according to the author [1], has 80.17% of accuracy. This model comes from a python package that is called pretrainedmodels [1]. In [4] the image content analysis was used to to build a probability distribution that made possible to classify an image according to its sentiment polarity, (negative, neutral or positive). So, with the InceptionResNetV2, a model was built that is fed with the union of the training and validation sets to increase the number of the images. The InceptionResNetV2 classified the contents of each dataset image into one of the ImageNet classes. Each of these images contains a sentiment classification in the training and validation sets that were used to build a table with the probability distribution of the image sentiment for each ImageNet class. This was then combined with text and image sentiment results to obtain the final classification.

3 Proposed Method

The method in [4] fused information from 3 different sources. In this paper we explore alternative network architectures that can improve the results obtained from the analysis of sentiment on isolated images.

3.1 Image Classification

The developed method for image analysis is based on a deep learning approach. This is implemented with Pytorch [6], which is a deep learning framework that supports several features and automatic differentiation. For this work, we explore three versions of the Resnet, which are, the ResNet18, the ResNet50, and the ResNet152. We explore too other architectures, the Inception V3 and DenseNet. We use these typologies because of they are the state-of-the-art methods that can reduce significantly the vanishing gradient problem. To use these models we need to set them up and prepare the data. To do that we follow the next steps.

3.2 Data Preprocessing

One of the biggest challenges in deep learning approaches is data quality. Any deep learning approach is hungry for data because it is through it that the network extracts and learns the features used for classification. The dataset used has many images with different scales and sizes. This fact can slow the training process. The pre-processing method used resizes each image to 224x224 in the case of the ResNets architectures, and 299x299 in the others.

This work was partially supported by Instituto de Telecomunicações under grant UID/EEA/50008/2019 and by project MOVES - Monitoring Virtual Crowds in Smart Cities (PTDC/EEI-AUT/28918/2017) financed by FCT - Fundação para a Ciência e a Tecnologia.