IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received April 15, 2017, accepted May 12, 2017, date of publication May 23, 2017, date of current version June 27, 2017.

Digital Object Identifier 10.1109/ACCESS.2017.2706674

Age Groups Classification in Social Network Using Deep Learning

RITA GEORGINA GUIMARÃES[1], RENATA L. ROSA[2], DENISE DE GAETANO[3], DEMÓSTENES Z. RODRÍGUEZ[1], (Senior Member, IEEE), AND GRAÇA BRESSAN[2]

[1]Federal University of Lavras, Lavras 37200-000, Brazil
[2]Polytechnic School, University of São Paulo, São Paulo 05508-010, Brazil
[3]University of Ulster, Coleraine BT52 1SA, Northern Ireland

Corresponding author: Rita Georgina Guimarães (rgg.rita@posgrad.ufla.br)

This work was supported in part by the University of São Paulo, in part by the Federal University of Lavras, and in part by the Minas Gerais State Agency for Research and Development (FAPEMIG).

ABSTRACT Social networks have a large amount of data available, but often, people do not provide some of their personal data, such as age, gender, and other demographics. Although the sentiment analysis uses such data to develop useful applications in people's daily lives, there are still failures in this type of analysis, either by the restricted number of words contained in the word dictionaries or because they do not consider the most diverse parameters that can influence the sentiments in a sentence; thus, more reliable results can be obtained, if the users profile information and their writing characteristics are considered. This research suggests that one of the most relevant parameter contained in the user profile is the age group, showing that there are typical behaviors among users of the same age group, specifically, when these users write about the same topic. A detailed analysis with 7000 sentences was performed to determine which characteristics are relevant, such as, the use of punctuation, number of characters, media sharing, topics, among others; and which ones can be disregarded for the age groups classification. Different learning machine algorithms are tested for the classification of the teenager and adult age group, and the deep convolutional neural network had the best performance, reaching a precision of 0.95 in the validation tests. Furthermore, in order to validate the usefulness of the proposed model for classifying age groups, it is implemented into the enhanced sentiment metric (eSM). In the performance validation, subjective tests are performed and the eSM with the proposed model reached a root mean square error and a Pearson correlation coefficient of 0.25 and 0.94, respectively, outperforming the eSM metric, when the age group information is not available.

INDEX TERMS Social network services, sentiment analysis, machine learning, text analysis, artificial neural networks, deep network.

I. INTRODUCTION

Nowadays, with the constant use of the Internet, users spend hours browsing on e-commerce sites, reading news about sports, journalism and entertainment, and expressing their opinions and sentiments in the form of comments on social networks about diverse topics. These comments can be analyzed to assess customer satisfaction that is a very useful information for service providers and product suppliers. Goldsmith et al. [1] explores the behavior of people that uses Internet for e-commerce and stresses the importance for evaluating the customer satisfaction in this type of services.

A tool capable of classifying sentiments and emotions should greatly facilitate the tasks of managers and behavioral analysis specialists. In general, there are several areas in which sentiment analysis can be applied, such as, business, marketing, beauty, fashion, sports, technology, health, among others. There are many applications already implemented that use sentiment analysis, for instances, detection of psychological diseases [2], detection of false profiles [3] to prevent criminals from attracting new victims [4], prediction the success or failure of a political candidacy, measure the spread of a disease, and determine the degree of criminality of a city [2].

Currently, there is a concern and a great effort to analyze data from online social networks to predict information that may reflect different aspects of the current reality [5]. The social network Twitter, due to its data availability policy, provides several short sentences, the tweets, which can be collected and analyzed. However, the informal and short sentences with many variations of language [3] do necessary the study of some parameters to improve the data analysis. Among them is the age that can directly influence the final sentiment of a sentence [6], [7]. Common characteristics, found during each phase of life, are taken into account in this type of analysis; specifically, those characteristics are clearly different in the teenager and adult age groups. It is important to note that in some social networks, the user age is not available either by the social network itself or even by the user for discretion reasons; as a consequence, the determination of a method to predict the user's age is relevant in the sentiment analysis.

There are few studies that take into account the influence of both age and gender in the manner a person can express a sentiment on weblogs [8]. Also, the difference in the style of use of the words, writing style, has already allowed predicting the age groups of blog authors with an accuracy of up to 80.32% [9], [10]. De Jonge et al. [11] investigated text-message abbreviations in high school and university students, such as only emoticons, slang, message length and spell errors. Huffaker et al. [12] examined the language use among teenagers on weblogs; they concluded that the most common language used was the acronyms and emoticons. Shapiro et al. [13] studied the frequency that the teenagers write on social networks. Each of these works treats about specific parameters; therefore, they do not reach a high accuracy.

In this context, the main contribution of this work is to demonstrate that parameters, such as, the use of punctuation which includes the emotion icons, the number of characters in the message or sentence lengths, slang, the use of Uniform Resource Locator (URL) to share media information, the number of people the user follows, the number of followers, the total number of tweets posted on social network and the approached topics are relevant to increase the assertiveness and accuracy for classifying the age group. Some of these parameters have already been used in other works, such as the use of slang, emotion icons and sentence lengths [14]–[16]; but they do not consider the parameters such as punctuation, URL, people the user follows, followers and total number of tweets. Each of these parameters was determined after a qualitative analysis performed manually and considering a huge amount of sentences collected from Twitter. Furthermore, our research also determined which parameters can be discarded at the time of classifying the age group, such as, citations with the use of the @ and hashtag symbols, and the sharing of messages. The educational level will not be considered in this work because they have already been tested [17] and presented low accuracy.

It is worth noting that different topics of context, such as health, family, politics, professional environment, among others, were considered. Also, it is important to note that although the study has been performed using the social network Twitter, it can be extended to other social networks because the parameters are usually the same.

In this research, in order to classify the teenager and adult age group, different machine learning algorithms were tested, and the Deep Convolutional Neural Network (DCNN) reached the best performance.

Furthermore, to validate the usefulness of the proposed model to classify age groups, it is implemented into the enhanced Sentiment Metric (eSM) [18]. In the performance validation, subjective tests are performed and the results are compared with the following sentiment metrics, Sentimeter-BR2 [19], [20], eSM without considering the proposed model, eSM with the proposed model, and eSM that considers the real age group. These experimental results demonstrated the relevance to consider the age group parameter, and the usefulness of the proposed model in the sentiment analysis.

The remainder of this paper is structured as follows. Section II presents some works related to the influence of the age group on the characteristics of the writing, an overview of sentiment analysis, and how machine learning can be used in this context. Section III introduces the proposed model for classifying age groups. Section IV presents the application of proposed model in a sentiment intensity metric. The results can be observed in Section V and then some discussions are presented in Section VI. Finally, the conclusions are presented in Section VII.

II. RELATED WORK

In this section, the main studies regarding to the influence of some parameters in the age group classification are treated. Also, the sentiment analysis and machine learning algorithms are discussed. In the sentiment analysis, some studies are cited to highlight that the user's age information is an important parameter to improve the performance of sentiment intensity metrics.

A. RELATIONSHIP BETWEEN THE AGE GROUP AND THE CHARACTERISTICS OF THE WRITING

The field of psychology shows the difference in behavior among people of different ages [21], [22]. Teenagers, in general, are not concerned about their privacy [23] and they post and disseminate a lot information on social networks; it could consider teenagers as persons up to eighteen, that is when they reach the age of majority, but for cultural reasons, some countries use the range between 13 and 19 years [21].

The age information is not always provided in some social networks, for instances, Twitter. After verifying that this information could actually alter the results of several analysis, some research [3], [17], [24] have worked on trying to predict it. One strategy used was to search for descriptions in the profile that contained the expressions "X years", "I have X years" or "I did X years", where X represents the user's age. However, it was verified that on Twitter, informing the age in the profile description is not a common habit [25] and therefore these studies would not provide reliable results.

It is common among teenager users of social networks to discuss more topics that occur in their daily lives, impacting their real world [10]. Topics such as relationships, school and friends are more frequent in this age group [12].

Adult users are concerned with their own images; then, they are more careful with the comments they write, and who can read them [26]; therefore, it is possible to find more sentences with positive sentiments, not using self-reference, making less use of negation [8], and consequently the use of slang also becomes less frequent [27].

The fact that adults write less about themselves can also be justified by the time users spend online; in adulthood, users have more commitments throughout the day, and teenagers spend more hours a day in online media; then, for teenagers, the social networks becomes the major means of expressing their opinions to the world [13].

In addition to the classic identity markers in adulthood, such as the topics of religion, ideology, politics, and work; adults are also characterized for using online media to express their comments. Adult users are accustomed to attach photos, videos or share links of another page that will complement the information that was initiated in the tweet [28].

This research will consider two main age groups, teenager and adult, because of the large difference in behavior between these two groups. Users of Twitter under the age of thirteen are not considered in this research, because many social networks require users to be at least thirteen years old to sign up in it. Thus, the teenager group is composed of users from 13 to 20 years old, known as teenagers and the adult group is composed by all users aged 20 or over.

B. SENTIMENT ANALYSIS

This section will cover studies that have been developed over the last years in order to achieve an automatic, reliable and realistic analysis of sentences extracted from the Internet. In this context, it is common to find studies that analyze the comments of e-commerce sites and also social networks by punctuating the sentiments of that comment. In this way, it is possible to gauge how much a product or service is well seen by the market or, likewise, it is possible to verify which aspects still need improvement. In addition, the sentiment analysis can be useful to analyze, almost in real time, a specific topic in order to determine some statistics.

There are many studies about sentiment analysis, but the majority does not consider the user profile such as the Sentimeter-Br2 metric that is based on a lexicon dictionary, in which each word has a positive or negative value of sentiment. This metric considers n-grams, adverbs and no stopwords, differentiates sentiment values depending on verbal tenses, in which verbs in past tense have a lesser sentiment value than verbs in the present tense. The Sentimeter-Br2 is based on the Sentimeter-Br [29].

The studies, explained in the following, addressed the impact of the user profile information in the sentiment analysis, with the aim of increasing the performance of the sentiment metrics.

1) The ANEW is a study that considered relevant to take into account the user information for the sentiment analysis [30]. It was studied how the presence or not of gender information would affect the final result.
2) The SentiWordNet is a sentiment metric used to classify the sentiment polarity automatically, with the aim of assisting the social media opinion analysis [31]. The authors analyzed how important is to consider the user profile, and later, the focus of the study was modified to cover the analysis different languages [32].
3) The eSM is another sentiment metric, which considered the characteristics of the user's profile, amongst them, the gender, level of education, geographic location and age are mentioned. The eSM is the association of a lexicon-based sentiment metric, the Sentimeter-Br2, with a correction factor based on the user's profile information. The eSM model of a sentence F_i is given by (1). Note that it is assumed that the age information is available in the user profile.

$$\begin{aligned} eSM(F_i) = {} & Sentimeter\_Br2(F_i) * C * exp(a_1 * A_1 \\ & + a_2 * A_2 + \ldots + a_n * A_n + g_1 * M \\ & + g_2 * F + e_1 * G + e_2 * nG + t1 * T2 \\ & + \ldots + tm * Tm) \end{aligned} \tag{1}$$

Where:

- C is a scale constant, obtained by subjective tests [18];
- $a_1 \ldots a_n$ are binary factors related to age groups $A_1 \ldots A_n$ are the weight factors of each age group, been considered four groups;
- g_1 and g_2 are binary factors related to the gender; M and F are the weight factors of gender, man or woman, respectively;
- e_1 and e_2 are binary factors related to educational level (higher education or not);
- G e nG are the weight factors of educational level, higher education or not, respectively.

Besides these works, others [33], [34], reported that in a rigorous sentiment analysis, it is important to have metrics for the detection of irony, since this can completely reverse the sentiment attributed to a sentence. In both works is stated that the user's age can influence how ironic the sentence is rated. However, in the Twitter, the age information is confidential and not mandatory.

Likewise, in [12] the authors showed that teenagers behave differently in the online environment, and it is possible to observe some peculiarities in the writing style, such as the topics that have an impact on their particular world.

The preference to post about topics that do not self-reference and deal with more positive information may be characteristic of adulthood users [26]; whereas slang use is more common in sentences posted by teenager users [27]. Moreover, the necessity to attach some media that represents the content being mentioned is also a characteristic of adulthood users [28].

As can be seen, the age parameter has been mentioned several times as a factor that influences the sentiment analysis and therefore, the techniques that improve the age classification are relevant. Thus, this work intends to group the characteristics mentioned previously [12], [27], [28] and others that are proposed in this research, such as the use of punctuations, abbreviations, symbols that express emotions and the characteristics of writing styles, in addition to others user's information such as tweets history, number of followers and number of people he or she follows.

C. MACHINE LEARNING

There are many machine learning models [35], which cover basic methods such as linear regression and tree models, as well as more sophisticated methods such as artificial neural networks or support vector machines. Typically, machine learning does not restrict data analysis to just one model, but compares many models, and chooses the one that achieves the best predictive accuracy. The machine learning area, also known as pattern recognition or data mining [36], is related to the extraction of patterns in large data sets. Frequently, the aim is to accurately predict a given response variable, for instance, age group, based on one or more preceding variables, such as writing characteristics.

Qualitative approaches are not usually considered in research on sentiment analysis [37], being necessary to filter or remove meaningless sentences that are considered as noise, mainly in corpus provided by Twitter [38]–[40], which contains abundant information. Many times, in the first phase of the data analysis, it is necessary to identify the main characteristics or patterns of the samples, and this task is performed manually by specialist. In this context, the research is carried out with large and filtered volumes of data, and spanning over a broad demographic, without considering specific people. Then, even messages with more personal information such as: "I was so homesick" or "I am going to start reducing carbohydrate", do not expose the users' private information in the results that will be obtained. After all, the goal is to draw patterns that exist in the manner of expressing themselves from each age group, and not just a few isolated cases.

In order to be able to work with the large amount of information and achieve the desired classification, the machine learning algorithms were used, which can provide results with high precision [41]; for instances, the algorithms based on decision trees (J48), support vectors (SMO) or artificial neural networks. The increasing use of Deep Leaning in various areas has been occurring in the recent years, such as image [42], [43] and speech recognition. The Deep Leaning algorithm allows computational models that are composed of multiple processing layers to learn representations of data with multiple levels of abstraction. Recently, Deep Learning methods have started to be also applied to text classification [44], [45], and the algorithms have obtained excellent results for classifying text models [46]–[49].

Deep Learning are usually interpreted in terms of the universal approximation theorem [50] or probabilistic inference [51]; the approximation theorem defines a class of universal approximators, which refers to the ability of neural networks of direct feeding with a single occult layer, of finite size, to approximate continuous functions. The probabilistic interpretation derives from the machine learning, which includes inference as well as optimization concepts such as training and tests, related to adaptation and generalization.

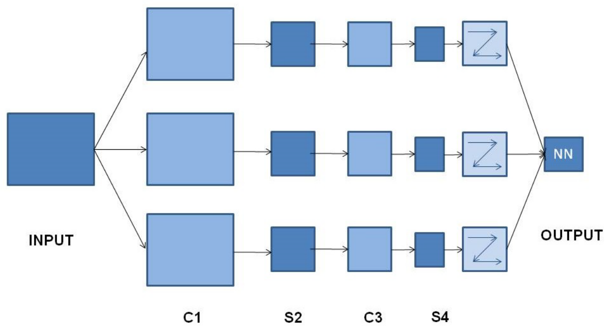

FIGURE 1. Convolutional neural network topology.

The DCNN can perform classification tasks, and it is composed of multiple layers, each one computes convolutional transforms [52]. Fig. 1 shows the topology of the DCNN, in which the C variables represent the convolution layers and the S variables represent the layers pool/sample; from C1 to S2 a convolution layer is present, from S2 to C3 a sub-sampling layer is present, from C3 to S4 exist another convolution layer, and from S4 to the output a fully connected Multilayer Perceptron (MLP) is represented.

As depicted in Fig. 1, it is possible to establish complex models, without necessarily hindering the interpretation of the results, besides being a resource widely used in the modeling of characteristics of human behavior [53]–[56].

III. PROPOSED MODEL FOR CLASSIFYING AGE GROUPS

This section presents the proposed model for classifying age groups that considers two phases, the data treatment extracted from social networks and the classification phase.

A. DATA TREATMENT EXTRACTED FROM SOCIAL NETWORKS

In order to obtain a more exact prediction of the age group, some information extracted directly from the social network was considered and some parameters that were considered important during the tests for this research. Among them is the punctuation mark, which was considered to know if the user had written some type of punctuation in the message; commas and end-point are disregarded because they are more common in any type of sentence. In this entry the symbols that express emotions, the called *emoticons*, were also considered as being punctuation. This last parameter has already been used in gender detection [12].

During the collection of the sample messages, we also noted that it is common among the teenagers to extend the spelling of some words, such as, "He is beautifullllllll !!!!!". In contrast, it is also possible to verify that the mean number of characters per word in sentences that have been written by teenagers is lower than the sentences written by adults, resulting in shorter tweets.

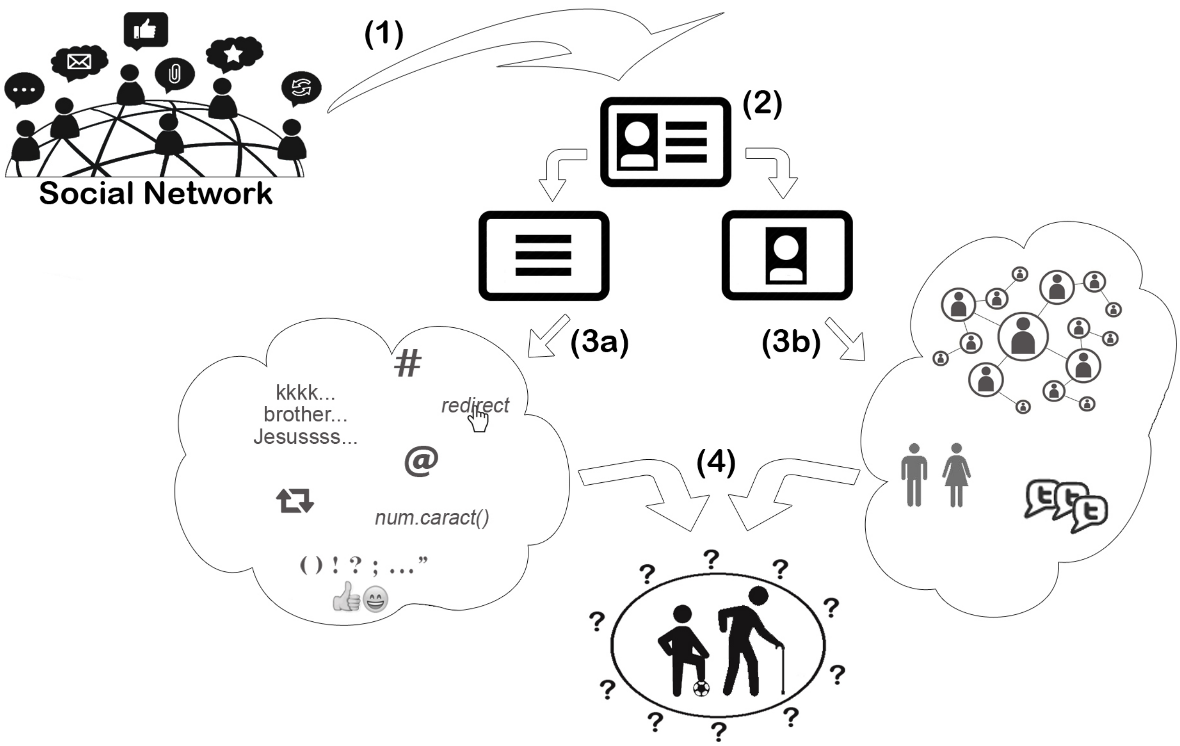

FIGURE 2. **Representation of the methodology for creating the proposed model.**

The abbreviations and variations in writing are considered in our research. The use of slang has been incremented with the predefined abbreviations in a dictionary, besides the spelling variations of the words. The entry that refers to the attached media, the URL is composed by the tweet messages that contain a link pointing to another page, or some kind of attachment such as photos or videos. We also considered if the message has the markers or symbols, "#" that highlight some topic, and "@" which is used to mention the name or nickname of another user.

Others entry parameters considered in this work are extracted directly from the user profile and they are part of his or her history on the social network. These are the number of people the user follows, the number of followers he or she owns, and the total number of tweets posted on his or her profile.

1000 sentences of each topic selected were used to classify users as teenagers or adults, the sentences were extracted via Twitter's Application Programming Interface (API). The keywords analyzed were: "medical agreement", "healthy eating", "carbohydrate", "pray", "professional work","mother" and "PEC 241"; in which the latest one is related to a political discussion in Brazil. The keywords used to collect the messages are intended to cover a large number of users, addressing different topics such as responsibilities, sports, health, politics, religion, work and family. The word dictionary used in this work considered different topics, in which each one has the respective words and their sentimental scores, depending on the topic.

A total number of 7000 sentences or samples was obtained, in which 60 contained identical information to other samples, characterizing as repeated messages; 660 were considered outliers because they presented very unusual characteristics in relation to the others, resulting in the 6280 valid samples that were used for analysis, in which 80% of samples were used as modeling and 20% were used as classifier. It is important to highlight that the analysis of the content of the samples was performed manually and all these samples own to users that their profile information, such as age and gender, are available.

Fig. 2 shows how each sample was analyzed. First, from one of the selected topic (1), the tweets were extracted from the social network containing the written message and user profile information (2); from the written sentence (3a) were analyzed some parameters and the others were obtained through the history and user profile (3b); the analysis of these parameters was performed using machine learning algorithms (4) to predict the age group that would best classify to each of the users.

In this research, the following parameters are considered to predict age group:

- rt: considers if the sentence is a retweet;
- arroba: this marker is represented by the symbol @;
- hashtag: this marker is represented by the symbol #;
- slang: the use of slang is considered and they are contained in the dictionary;
- punctuation: The punctuation considers characters such as " " ? ! ... : ( ) / and the symbols that express emotions;

- url: the message can have a link pointing to another page or some kind of attachment like photos or videos, as a media sharing, which is represented by an URL;
- characters: it is a numeric parameter that represents the amount of characters on the sentence, obeying Twitter's counting policy;
- follow: it is a numeric parameter that represents the number of users that a specific user follows;
- followers: it is a numeric parameter that represents the number of followers each user owns;
- tweets: it is a numeric parameter that represents the total tweets the user writes on his or her profile;
- topic: the main topic of the sentence is considered;
- gender: the gender of the user who writes the message, which is represented by male and female genders;
- teenager: the parameter that represents the output of the machine learning algorithm, it is represented as teenager or no teenager (adult).

B. CLASSIFICATION PHASE

Once defined the most relevant parameters to predict the users' age group, the machine learning algorithms were used; in which each of the parameters is an input of the algorithms.

The normalization of data considers the parameter types, such as binary or numeric. The parameters as rt, @, hashtag, slang, punctuation, URL and the definition whether the user is teenager or not are binary, because they have only the YES or NO response, if the answer is positive or negative, respectively. Also, the gender parameter is binary, in which the symbol F was assigned for woman, in the gender field, and M for man. The other entries: characters, follow, followers, tweets and topic are numeric parameters that represent the actual extracted value as seen in Fig. 3.

```
@relation twitter

@attribute rt {YES,NO}
@attribute arroba {YES,NO}
@attribute hashtag {YES,NO}
@attribute slang {YES,NO}
@attribute punctuation {YES,NO}
@attribute url {YES,NO}
@attribute characters numeric
@attribute follow numeric
@attribute followers numeric
@attribute tweets numeric
@attribute topic numeric
@attribute gender {F,M}
@attribute teenager {YES,NO}

@data

NO,YES,NO,NO,NO,NO,116,310,947,21876,5,M,NO
NO,YES,NO,NO,YES,NO,130,193,11,56,5,M,NO
NO,NO,NO,NO,NO,NO,140,509,4562,84457,5,M,NO
NO,NO,NO,NO,NO,NO,88,221,136,9116,1,F,YES
NO,NO,NO,NO,NO,NO,33,286,823,36194,1,F,YES
NO,NO,NO,YES,YES,NO,132,62,124,1488,1,F,YES
```

FIGURE 3. Representation of normalized data.

In the machine learning, the tests were performed using the following algorithms, Artificial Neural Networks, Decision Trees, Random Forest and Support Vector Machine (SVM).

Artificial Neural Networks have a large generalization capability, and can approximate functions used for both regression and classification. The principle of this network is the propagation of information through the artificial neurons, using the synaptic weights to modify the input signal and generate an output signal. The learning of synaptic weights is done by adjusting the propagation of the error in the reverse direction [57].

The tests were implemented using both, the Waikato Environment for Knowledge Analysis (WEKA) that is a very popular in similar research, and a code developed in Python programming language to run Deep Learning algorithms [45].

It is worth noting that the recursive neural network used in some studies [46] obtained good results in terms of constructing sentence representations, but it builds a tree structure and in a sentence the later words have more dominant value than earlier words; this can reduce the effectiveness to capture the semantics of a whole sentence. Instead, the DCNN captures better the semantic of texts compared to the recursive neural network [58]. For this reason, in this work the DCNN was used.

The DCNN used in the work needed the following steps to train the deep learning model for classification:

- it used the word2vec neural language model;
- a convolutional neural network was performed to refine the supervised corpus;
- by the end, parameters of the network were used to initialize the network obtained in a previous stage to training the supervised corpus.

The machine learning tool, used in this work, allows adjustment of all neural network parameters, such as, learning rate, momentum rate, number of times, training method, number of layers and neurons, among other parameters [59].

In the experiments, the topology of DCNN used a linear learning rate of 0.0004, batch size of 128, decay factor of 0.001 and momentum of 0.9. Each network of the algorithm was trained using 100 epochs. The Softmax classifier was used in the DCNN.

As stated before, the tests were also performed using decision tree algorithm that is common in the context of sentiment analysis [60]. The principle of the decision tree is to group instances in a recursive way, minimizing the variability of the classes. To do this, the algorithm uses the values of the attributes in each nonleaf node and places the instances on the leaf nodes. Thus, each internal node corresponds to a classification decision [56]. The tree algorithm is not able to perform regression and it plays only the role of classifier.

In the test performed by the Random Forest algorithm, all trees are trained using the same parameters, but in different training sets. These sets are generated from the original training set using the initialization procedure: for each training set the same number of vectors as in the original set are chosen randomly. It may occur that vectors are used more than once and others are absent [61], not all variables are used to find the best separation, but a random subset of them.

The classification with Support Vector Machine (SVM), which is also used in machine learning proposals [62] has been widely used in tasks for emotion classification [54] with good generalization properties [56]. SVM has discriminatory methods that learn boundaries between classes [63], performing a binary classification based on the separation of hyperplanes; a separator is chosen in order to maximize the distances of these hyperplanes and the nearest formation vectors, which are called support vectors [64]. Unlike other techniques, the functions of the probability model do not need be known previously [65]. This is very important for generalization purposes, because in practical situations there may not be enough information about these functions and distributions between inputs and outputs.

In the experimental test carried out in this work, the same samples were used for all machine learning algorithms to evaluate the relevance of each parameter.

Later, the model generated by machine learning was applied in the sentences extracted from the social network, the age group is classified and each sentence is used in the eSM sentiment metric to evaluate the performance of the proposed model.

IV. APPLICATION OF THE PROPOSED MODEL IN A SENTIMENT INTENSITY METRIC

The validation of the proposed model for classifying age groups is presented in this section.

The eSM sentiment metric was used to evaluate the usefulness of the proposed model for age groups classification. The eSM works with user profile, including the age group that is provided, in case the social network offers this data and/or the user gives permission. The eSM does not reach its best performance if the age group is not available; then, the proposed model of classifying age groups is used to obtain the information, simulating the missing of the age group information, classifying the user in the teenager or adult age group. It is worth noting that the eSM uses four age groups, therefore, the eSM model was modified to accept only two groups, teenager and adult. Also, the Sentimeter-Br2 metric was also used in the performance tests; this metric does not considers any user's profile information, then it will be used to demonstrate the relevance of the profile information in the accuracy of sentiment intensity determination.

The subjective tests were performed in two phases. In the first phase, the test were carried out in a laboratory environment to obtain the correct parameters of the user's profile; in a second phase, remote subjective tests were performed to extract more sentences from a social network.

A. SUBJECTIVE TESTS IN A LABORATORY ENVIRONMENT

The subjective tests were performed in a laboratory environment to obtain the correct parameters of the user's profile. These tests approached a wider diversity of assessor profiles, since the region of birth (North, Northeast, Midwest, Southeast and South of Brazil), age, educational level and gender. Furthermore, the participants filled in some information, such as his or her user name in the social network, Twitter, which was recorded in a database for continuing monitoring the assessors remotely. Later, the assessors began to write sentences in the social network and scored his or her own sentences based on a sentiment intensity scale, from -5 to $+5$, in which -5 and $+5$ represented the most negative and the most positive values, respectively. To perform these tests, assessors used a mobile device with an application installed; the first screen shot of that application is presented in Fig. 4. This application is used to store the user profile information, such as, age and gender among others.

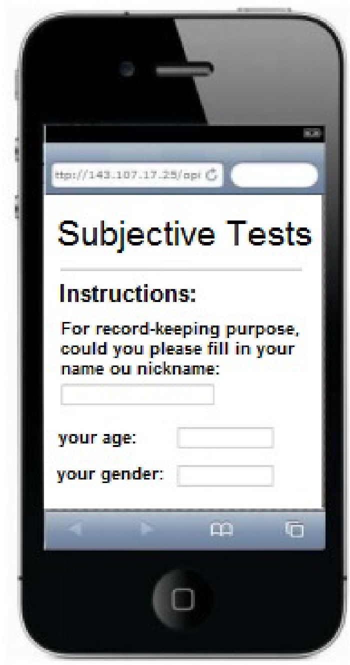

FIGURE 4. Application tests in a mobile device.

The assessors' information were collected in a database and the field corresponding to age was separated into two groups, teenager and adult, with users aged 20 or over were called adults.

A total number of 76 assessors participated of tests, consisting of 38 women and 38 men. Later, each assessor posted sentences remotely on the social network Twitter, which sentences were captured by a script. The sentences were analyzed by the same assessor who posted the sentences and by the Sentimeter-Br2, the eSM metric with and without the proposed model to predict the age group. In this phase, each assessor wrote only 14 sentences, two sentences for each topic, reaching a total of 1064 sentences. The average time spent by each user to attend the test instructions, complete the information required and the write the sentences was approximately 42 minutes.

B. SUBJECTIVE TESTS MONITORED REMOTELY

The assessors that participated in the face-to-face tests were monitored remotely over a test period of 2 weeks, every hour. In total, 5,323 sentences were extracted from the social network. It is important to note that assessors authorized to be monitored, and they also provided the necessary social network information. In these tests, the number of messages written by each assessor was not limited, and they also provided their real sentiment for each sentence.

The purpose of the tests was to calculate the value of the sentiment intensity based on the eSM metric that was fed with real data of the user profile and with the Sentimeter-Br2 metric that does not consider the user profile.

The subjective tests were performed to extract the real value of the age groups and the real sentiments scored by the assessors. Thus, the proposed model for classifying age groups could be applied in the collected sentences, wherein the real age group was deleted and replaced by the age group generated by the proposed model.

The performance results in the validation phase consider the following statistical functions, the root mean square error (RMSE), the maximum and average error and the Pearson Correlation factor (PCC), all of them based on the sentiment predicted values by the sentiment metrics and the values scored by the assessors.

V. RESULTS

In this section will be presented the results of the machine learning classification for generating the proposed model, and the validation of the proposed model in conjunction with the eSM trough subjective tests.

A. MACHINE LEARNING FOR CLASSIFYING MODEL

As stated before, 7000 sentences were collected from social networks, and 6280 sentences were considered as valid and they were used to determine the age group classification model. From this number, 80% of sentences (5024) were used to training the model, and the 20% (1256) were used to evaluate the performance of the model.

In the model training phase, the DCNN obtained a precision of 0.910 for users classified as teenagers and 0.956 for users classified as adults with the methodology used; in addition, the recall values reached 0.919 and 0.931, respectively. By means of this data, it is possible to calculate the F-Measure which is the harmonic average between the information obtained. In this research, the F-Measure reached the value of 0.930 for the general classification, as can be seen in Table 1 together with the other results obtained by the other machine learning algorithms.

In order to clarify the relevance of some parameters, Fig. 5 partially demonstrates the behavior of the test performed by the decision tree algorithm, J48, which permits see the parameters and their respective weights.

Observing the weight assigned to each of the parameters by the Neural Network and also the positioning of the nodes by the Decision Tree, we verified that some characters such as the markers "@" and "#", and also the parameter known as retweet (rt), could be withdrawn without losses accuracy in the final classification. The results obtained in the second test, using the same structures and previous samples, can be seen in Table 2.

TABLE 1. Machine learning results to classifying age groups considering all parameters - training phase.

Algorithm	Precision	Recall	F-Measure	Class
Multilayer Perceptron	0.799	0.814	0.810	Teenager
	0.867	0.854	0.860	Adult
	0.836	0.836	**0.840**	**General**
DCNN	0.910	0.919	0.910	Teenager
	0.956	0.931	0.940	Adult
	0.938	0.916	**0.930**	**General**
Decision Tree	0.792	0.811	0.800	Teenager
	0.858	0.841	0.850	Adult
	0.829	0.830	**0.830**	**General**
Random Forest	0.792	0.863	0.830	Teenager
	0.891	0.837	0.860	Adult
	0.849	0.849	**0.850**	**General**
SVM	0.810	0.819	0.810	Teenager
	0.869	0.849	0.860	Adult
	0.840	0.838	**0.840**	**General**

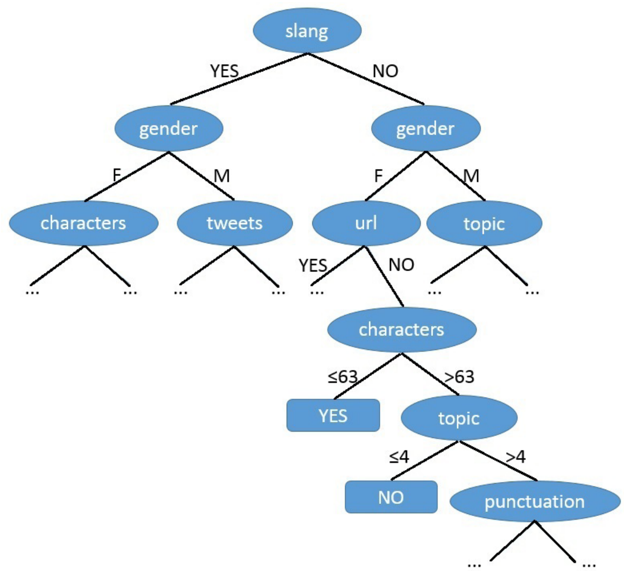

FIGURE 5. Representation of the decision tree.

TABLE 2. Machine learning results to classify age groups after discard no relevant parameters - training phase.

Metric	Precision	Recall	F-Measure	Class
Multilayer Perceptron	0.819	0.838	0.830	Teenager
	0.879	0.862	0.871	Adult
	0.882	0.869	**0.880**	**General**
DCNN	0.929	0.936	0.930	Teenager
	0.967	0.965	0.970	Adult
	0.952	0.951	**0.950**	**General**
Decision Tree	0.822	0.818	0.820	Teenager
	0.871	0.873	0.870	Adult
	0.842	0.851	**0.850**	**General**
Random Forest	0.798	0.869	0.830	Teenager
	0.893	0.847	0.870	Adult
	0.854	0.853	**0.850**	**General**
SVM	0.816	0.814	0.810	Teenager
	0.869	0.871	0.870	Adult
	0.851	0.850	**0.850**	**General**

TABLE 3. Machine learning results to classify age groups - validation phase.

Metric	Precision	Recall	F-Measure	Class
Multilayer Perceptron	0.801	0.829	0.810	Teenager
	0.873	0.854	0.860	Adult
	0.848	0.849	**0.850**	**General**
DCNN	0.903	0.912	0.910	Teenager
	0.956	0.958	0.960	Adult
	0.937	0.937	**0.940**	**General**
Decision Tree	0.812	0.813	0.810	Teenager
	0.864	0.863	0.860	Adult
	0.833	0.842	**0.840**	**General**
Random Forest	0.784	0.858	0.820	Teenager
	0.888	0.841	0.860	Adult
	0.845	0.849	**0.850**	**General**
SVM	0.800	0.809	0.800	Teenager
	0.853	0.866	0.860	Adult
	0.833	0.843	**0.840**	**General**

As can be observed from Table 2, the parameters "@", "#" and retweet did not influence much in the classification.

After the network training phase was concluded, the model can be used to test any other samples with the weights and corrected rates. Then, the 1256 remaining samples were used to evaluate the performance of the model. Table 3 shows the results obtained in the validation phase, in which can be observed that the DCNN reached a general F-Measure value of 0.94 that is very similar to the F-Measure value reached in the training phase.

B. EVALUATION OF THE USEFULNESS OF THE PROPOSED MODEL IN SENTIMENT INTENSITY METRICS

A total number of 6387 sentences were used on the evaluation of the usefulness of the proposed model in sentiment intensity metrics, 1064 sentences were collected in the face-to-face tests performed in the laboratory and 5323 sentences collected in the remote tests.

TABLE 4. Performance assessment of the sentiment metric that considers the proposed model to predict age group.

	Sentimeter-Br2	eSM without proposed model	eSM with proposed model	eSM with real age group
RMSE	0.34	0.29	0.25	0.23
Maximum error	0.26	0.19	0.15	0.12
Average error	0.19	0.15	0.10	0.06
PCC	0.88	0.90	0.94	0.96

Table 4 presents the performance assessment and the usefulness of the proposed model considering the RMSE, Maximum Error, Average Error and PCC. This evaluation considered the following scenarios: the Sentimeter-Br2, which do not consider any profile parameter; the eSM without both the age information and the proposed model; the eSM considering the age group classified by the proposed model; and finally, the eSM considering the real age group filled by assessors from subjective face-to-face tests. The sentiment intensity, scored by the assessors, represents the ground-truth values.

As can be observed in Table 4, the Sentimeter-Br2 presents the worst performance because it is a sentiment metric that does not take into consideration any data from the user's profile, and as can be expected, the eSM with the real age group (four groups) reached the best performance. Here, the most important comparison is between the eSM with and without the proposed model, in which the proposed model helped to get a better performance. This is very important, because in many cases the users' age are not available in social networks, and the prediction using the proposed model improve the accuracy of a sentiment intensity metric.

VI. DISCUSSIONS

The gender has already been considered as an useful information to predict the age of the users [66], and as can be seen in the representation of Fig. 5, it occupies one of the nodes higher up the tree; this means that gender information has a great importance in the performance of sentiment metrics. Although studies [25] show that the criteria for gender classification such as name distinction or word frequency are not efficient, others were able to predict the gender with an accuracy of approximately 80% [67], analyzing the frequency of certain words. Efforts to automate this process are valid in order to be able to include it in sentiment metrics. In this research we focus on the prediction of the age information because this information sometimes is not available, and the analysis of its impact on the precision of a sentiment metric is also studied.

Our experimental results demonstrated that the URL attribute, which refers to the user mentioning other pages or attaching photos and videos to the tweet, is a relevant feature to predict the user' age group. This means that it will be more common to find adults commenting on news or even advertisements on their profile, differentiating them from teenagers.

Attributes such as the use of slang and the number of characters in the message are also very specific in each age group and because they are easily detected, they guarantee greater reliability to the results.

According to the results, we verified that there are topics over which adults or teenagers express themselves more frequently. The topic parameter was also of great importance because the keywords were previously defined, thus, in addition to reaching a wide audience, they still balanced the amount of samples in each age group. It is clear that in certain topics it is possible to predict who will be the target audience, facilitating the classification and proving the effectiveness of the proposed age group classification model.

Also, the parameters removed did not affect the performance of the proposed model; then, the processing of this information is not necessary, decreasing the time consumed for the analysis to predict age groups.

The most well accepted machine learning algorithms were used, and DCNN obtained the best results reaching a F-Measure value of 0.940 in the validation phase of the age group classification model, which is a very reliable result.

The usefulness of the proposed model was demonstrated. In the cases that the user age information is not available, the age prediction model helps to improve the accuracy of sentiment intensity metrics. Moreover, the proposed model can be used in Recommendation Systems in order to suggest the most suitable content for each user. Although it is reported that there may be differences between the so-called ''social age'' and ''biological'' age [24], in which the goal is to know in which classification each user would identify better.

Finally, it is important to stress that the eSM metric considers four age groups and the proposed age classification model considers two age groups. However, the results presented in Table 4 shows that the performance of the ''eSM with proposed model'' and ''eSM with real age group'' reached similar PCC and RMSE values. This fact demonstrated that the teenager and adult groups are the most relevant age groups to be considered.

VII. CONCLUSION

In order to obtain the most relevant parameters, an extensive number of sentences (7000) were analyzed qualitatively, to determine the characteristics of teenager and adults age groups, considering the writing style and both users' history and profile. The experimental results show that the parameters used in this research can reach a high accuracy for determining the age groups of Twitter users.

Some parameters have been removed because they do not influence the final classification result, make it clear they should not be considered or applied.

The DCNN was the machine learning algorithm that reached the best results for age groups classification.

The importance to consider data users' profile in a sentiment intensity metric is claimed in several studies. In this work, the Sentimeter-Br2 presented the worst performance compared with the eSM metric that considered the data user profile.

Social Networks do not always provide users information, or users restrict their personal information. In these cases, the proposed model to predict age group is very important to improve the performance of sentiment intensity metrics. In this research, the eSM metric was used, and the scenarios in which the age information is and is not available were tested. The scenarios in which the age is not available was improved by our proposed model. Furthermore, the proposed model is able to work with other sentiment metrics.

Results presented in Table 4 demonstrated that only considering teenagers and adults age groups, sentiment metrics can obtain similar results in relation to the cases in which real age is considered.

ACKNOWLEDGMENT

The authors wish to thank the University of São Paulo, the Engineering Department and the Computer Science Department at the Federal University of Lavras for the motivation at the research in social network area and sentiment analysis.

REFERENCES

[1] R. E. Goldsmith, ''Explaining and predicting consumer intention to purchase over the internet: An exploratory study,'' *J. Marketing Theory Pract.*, vol. 10, no. 2, pp. 22–28, Mar. 2002.
[2] R. G. Guimarães, D. Z. Rodríguez, R. L. Rosa, and G. Bressan, ''Recommendation system using sentiment analysis considering the polarity of the adverb,'' in *Proc. IEEE Int. Symp. Consum. Electron. (ISCE)*, Sao Paulo, Brazil, Sep. 2016, pp. 71–72.
[3] C. Peersman, W. Daelemans, and L. Van Vaerenbergh, ''Predicting age and gender in online social networks,'' in *Proc. 3rd Int. Workshop Search Mining User-Generated Contents*, Glasgow, Scotland, Oct. 2011, pp. 37–44. [Online]. Available: http://doi.acm.org/10.1145/2065023.2065035
[4] J. van de Loo, G. De Pauw, and W. Daelemans, ''Text-based age and gender prediction for online safety monitoring,'' *Comput. Linguistics Netherlands*, vol. 5, no. 1, pp. 46–60, Dec. 2016.
[5] R. M. Filho, J. M. Almeida, and G. L. Pappa, ''Twitter population sample bias and its impact on predictive outcomes: A case study on elections,'' in *Proc. IEEE/ACM Int. Conf. Adv. Social Netw. Anal. Mining*, Paris, France, Aug. 2015, pp. 1254–1261.
[6] D.-P. Nguyen, R. Gravel, R. Trieschnigg, and T. Meder, '''How old do you think I am?' A study of language and age in Twitter,'' in *Proc. 7th Int. AAAI Conf. Weblogs Social Media*, Palo Alto, CA, USA, Jul. 2013, pp. 439–448.
[7] L. Sloan, J. Morgan, P. Burnap, and M. Williams, ''Who tweets? Deriving the demographic characteristics of age, occupation and social class from twitter user meta-data,'' *PLoS ONE*, vol. 10, no. 3, pp. 1–20, Mar. 2015.
[8] J. Schler, M. Koppel, S. Argamon, and J. W. Pennebaker, ''Effects of age and gender on blogging,'' in *Proc. AAAI Spring Symp., Comput. Approaches Anal. Weblogs*, Stanford, CA, USA, Mar. 2006, pp. 199–205.
[9] S. Goswami, S. Sarkar, and M. Rustagi, ''Stylometric analysis of bloggers' age and gender,'' in *Proc. Int. AAAI Conf. Web Social Media*, San Jose, CA, USA, May 2009, pp. 214–217.
[10] S. Argamon, M. Koppel, J. W. Pennebaker, and J. Schler, ''Mining the blogosphere: Age, gender and the varieties of self-expression,'' *First Monday*, vol. 12, no. 9, pp. 214–217, May 2007.
[11] S. De Jonge and N. Kemp, ''Text-message abbreviations and language skills in high school and university students,'' *J. Res. Reading*, vol. 35, no. 1, pp. 49–68, Oct. 2010.
[12] D. A. Huffaker and S. L. Calvert, ''Gender, identity, and language use in teenage blogs,'' *J. Comput.-Mediated Commun.*, vol. 10, no. 2, pp. 1–24, Jun. 2005.
[13] L. A. S. Shapiro and G. Margolin, ''Growing up wired: Social networking sites and adolescent psychosocial development,'' *Clin. Child Family Psychol. Rev.*, vol. 17, no. 1, pp. 1–18, Mar. 2014.
[14] S. Rosenthal and K. McKeown, ''Age prediction in blogs: A study of style, content, and online behavior in pre-and post-social media generations,'' in *Proc. 49th Annu. Meeting Assoc. Comput. Linguistics*, Portland, OR, USA, Jun. 2011, pp. 763–772.
[15] D. Rao, D. Yarowsky, A. Shreevats, and M. Gupta, ''Classifying latent user attributes in twitter,'' in *Proc. Int. Workshop Search Mining User-Generat. Contents*, Toronto, ON, Canada, Oct. 2010, pp. 37–44.
[16] F. Barbieri, ''Patterns of age-based linguistic variation in American English,'' *J. Sociolinguistics*, vol. 12, no. 1, pp. 58–88, Jan. 2008.
[17] L. Zheng, K. Yang, Y. Yu, and P. Jin, ''Predicting age range of users over microblog dataset,'' *Int. J. Database Theory Appl.*, vol. 6, no. 6, pp. 85–94, Oct. 2013.
[18] R. L. Rosa, D. Z. Rodriguez, and G. Bressan, ''Music recommendation system based on user's sentiments extracted from social networks,'' *IEEE Trans. Consum. Electron.*, vol. 61, no. 3, pp. 359–367, Oct. 2015.
[19] R. L. Rosa, D. Z. Rodriguez, and G. Bressan, ''SentiMeter-Br: Facebook and twitter analysis tool to discover consumers sentiment,'' in *Proc. 9th Adv. Int. Conf. Telecommun.*, Rome, Italy, Jan. 2013, pp. 61–66.
[20] R. L. Rosa, D. Z. Rodríguez, and G. Bressan, ''SentiMeter-Br: A social Web analysis tool to discover consumers' sentiment,'' in *Proc. IEEE 14th Int. Conf. Mobile Data Manage.*, Milan, Italy, Jun. 2013, pp. 122–124.
[21] S. M. Sawyer *et al.*, ''Adolescence: A foundation for future health,'' *Lancet*, vol. 379, no. 9826, pp. 1630–1640, Apr. 2012.
[22] J. Y. Jang, K. Han, P. C. Shih, and D. Lee, ''Generation like: Comparative characteristics in instagram,'' in *Proc. 33rd Annu. ACM Conf. Human Factors Comput. Syst.*, Seoul, South Korea, Apr. 2015, pp. 4039–4042.
[23] S. Utz and N. C. Krämer, ''The privacy paradox on social network sites revisited: The role of individual characteristics and group norms,'' *J. Psychosoc. Res. Cyberspace*, vol. 3, no. 2, pp. 73–79, Nov. 2009.

[24] D.-P. Nguyen *et al.*, "Why gender and age prediction from tweets is hard: Lessons from a crowdsourcing experiment," in *Proc. 25th Int. Conf. Comput. Linguistics*, Dublin, Ireland, Aug. 2014, pp. 1950–1961.

[25] J. A. B. L. Filho, R. Pasti, and L. N. de Castro, "Gender classification of twitter data based on textual meta-attributes extraction," *Adv. Intell. Syst. Comput.*, vol. 444, pp. 1025–1034, Mar. 2016.

[26] J. W. Pennebaker and L. D. Stone, "Words of wisdom: Language use over the life span," *J. Personality Social Psychol.*, vol. 85, no. 2, p. 291, Aug. 2003.

[27] H. A. Schwartz *et al.*, "Personality, gender, and age in the language of social media: The open-vocabulary approach," *PLoS ONE*, vol. 8, no. 9, pp. 73–79, Nov. 2013.

[28] T. A. Pempek, Y. A. Yermolayeva, and S. L. Calvert, "College students' social networking experiences on Facebook," *J. Appl. Develop. Psychol.*, vol. 30, no. 3, pp. 227–238, Jan. 2009.

[29] R. L. Rosa, D. Z. Rodríguez, and G. Bressan, "SentiMeter-Br: A new social Web analysis metric to discover consumers' sentiment," in *Proc. IEEE Int. Symp. Consum. Electron. (ISCE)*, Las Vegas, NV, USA, Jan. 2013, pp. 153–154.

[30] M. M. Bradley and P. J. Lang, "Affective norms for English words (ANEW): Instruction manual and affective ratings," Center Res. Psychophysiol., Univ. Florida, Gainesville, FL, USA, Tech. Rep. C-1, 1999.

[31] A. Esuli and F. Sebastiani, "SentiWordNet: A publicly available lexical resource for opinion mining," in *Proc. Int. Conf. Lang. Resour. Eval.*, vol. 6. Genoa, Italy, May 2006, pp. 417–422.

[32] K. Denecke, "Using sentiwordnet for multilingual sentiment analysis," in *Proc. IEEE 24th Int. Conf. Data Eng. Workshop*, Cancun, Mexico, Apr. 2008, pp. 507–512.

[33] C. Bosco, V. Patti, and A. Bolioli, "Developing corpora for sentiment analysis: The case of irony and senti-TUT," *IEEE Intell. Syst.*, vol. 28, no. 2, pp. 55–63, Mar. 2013.

[34] D. Irazú, H. Farías, V. Patti, and P. Rosso, "Irony detection in Twitter: The role of affective content," *ACM Trans. Internet Technol.*, vol. 16, no. 3, pp. 1–24, Jul. 2016.

[35] Z. J. Lu, "The elements of statistical learning: Data mining, inference, and prediction," *J. Roy. Statist. Soc. A*, vol. 173, no. 3, pp. 693–694, Feb. 2009.

[36] B. Meuleman and K. Scherer, "Nonlinear appraisal modeling: An application of machine learning to the study of emotion production," *IEEE Trans. Affect. Comput.*, vol. 4, no. 4, pp. 398–411, Oct. 2013.

[37] F. Ren and K. Matsumoto, "Semi-automatic creation of youth slang corpus and its application to affective computing," *IEEE Trans. Affect. Comput.*, vol. 7, no. 2, pp. 176–189, Apr. 2016.

[38] A. Fraisse and P. Paroubek, "Toward a unifying model for opinion, sentiment and emotion information extraction," in *Proc. Int. Conf. Lang. Resour. Eval.*, Reykjavik, Iceland, May 2014, pp. 3881–3886.

[39] A. Qadir and E. Riloff, "Bootstrapped learning of emotion hashtags #hashtags4you," in *Proc. 4th Workshop Comput. Approaches Subjectivity, Sentiment Social Media Anal.*, Atlanta, GA, USA, Nov. 2013, pp. 2–11.

[40] A. Fraisse and Paroubek, "Twitter as a comparable corpus to build multilingual affective lexicons," in *Proc. Workshop Build. Using Comparable Corpora*, Reykjavik, Iceland, May 2014, pp. 26–31.

[41] A. Neviarouskaya and M. Aono, "Sentiment word relations with affect,judgment, and appreciation," *IEEE Trans. Affect. Comput.*, vol. 4, no. 4, pp. 425–438, Oct./Dec. 2013.

[42] I. J. Goodfellow, Y. Bulatov, J. Ibarz, S. Arnoud, and V. Shet, "Multi-digit number recognition from street view imagery using deep convolutional neural networks," *Comput. Vis. Pattern Recognit.*, vol. 6, no. 2, pp. 1–13, May 2015.

[43] H.-C. Shin, M. R. Orton, D. J. Collins, S. J. Doran, and M. O. Leach, "Stacked autoencoders for unsupervised feature learning and multiple organ detection in a pilot study using 4D patient data," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 35, no. 8, pp. 1930–1943, Aug. 2013.

[44] T. Chen, R. Xu, Y. He, and X. Wang, "Improving sentiment analysis via sentence type classification using BiLSTM-CRF and CNN," *Expert Syst. Appl.*, vol. 72, pp. 221–230, Nov. 2017.

[45] Y. Kim, "Convolutional neural networks for sentence classification," in *Proc. Conf. Empirical Methods Natural Language Process.*, Doha, Qatar, Oct. 2014, pp. 1746–1751.

[46] R. Socher *et al.*, "Recursive deep models for semantic compositionality over a sentiment treebank," in *Proc. Conf. Empirical Methods Natural Lang. Process.*, Seattle, WA, USA, Oct. 2013, pp. 1631–1642.

[47] X. Zheng, H. Chen, and T. Xu, "Deep learning for Chinese word segmentation and POS tagging," in *Proc. Conf. Empirical Methods Natural Lang. Process.*, Seattle, WA, USA, Jun. 2013, pp. 647–657.

[48] X. Glorot, A. Bordes, and Y. Bengio, "Domain adaptation for large-scale sentiment classification: A deep learning approach," in *Proc. 28th Int. Conf. Mach. Learn.*, Washington, DC, USA, Jul. 2011, pp. 513–520.

[49] X. Zhang, J. Zhao, and Y. LeCun, "Character-level convolutional networks for text classification," in *Proc. 28th Int. Conf. Neural Inf. Process. Syst.*, Montreal, QC, Canada, Apr. 2015, pp. 649–657.

[50] K. Hornik, M. Stinchcombe, and H. White, "Multilayer feedforward networks are universal approximators," *Neural Netw.*, vol. 2, no. 5, pp. 359–366, 1989.

[51] K. P. Murphy, *Machine Learning: A Probabilistic Perspective*. Cambridge, MA, USA: MIT Press, 2013.

[52] D. S. Touretzky, Ed., *Advances in Neural Information Processing Systems*, vol. 2. San Francisco, CA, USA: Morgan Kaufmann, 1990.

[53] E. Delaherche, M. Chetouani, A. Mahdhaoui, C. Saint-Georges, S. Viaux, and D. Cohen, "Interpersonal synchrony: A survey of evaluation methods across disciplines," *IEEE Trans. Affect. Comput.*, vol. 3, no. 3, pp. 349–365, Jul. 2012.

[54] Z. Zeng, M. Pantic, G. I. Roisman, and T. S. Huang, "A survey of affect recognition methods: Audio, visual, and spontaneous expressions," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 31, no. 1, pp. 39–58, Jan. 2009.

[55] M. Pantic and L. J. M. Rothkrantz, "Toward an affect-sensitive multimodal human-computer interaction," *Proc. IEEE*, vol. 91, no. 9, pp. 1370–1390, Sep. 2003.

[56] B. Schuller, A. Batliner, S. Steidl, and D. Seppi, "Recognising realistic emotions and affect in speech: State of the art and lessons learnt from the first challenge," *Speech Commun.*, vol. 53, nos. 9–10, pp. 1062–1087, Jan. 2011.

[57] C. M. Bishop, *Neural Networks for Pattern Recognition*. New York, NY, USA: Oxford Univ. Press, 1995.

[58] S. Lai, L. Xu, K. Liu, and J. Zhao, "Recurrent convolutional neural networks for text classification," in *Proc. 29th AAAI Conf. Artif. Intell.*, Austin, TX, USA, Jan. 2015, pp. 2267–2273.

[59] J. Deriu, M. Gonzenbach, F. Uzdilli, A. Lucchi, V. De Luca, and M. Jaggi, "Sentiment classification using an ensemble of convolutional neural networks with distant supervision," in *Proc. Semantic Eval.*, San Diego, CA, USA, Jun. 2016, pp. 1124–1128.

[60] G. Briones, K. Amarasinghe, and B. T. McInnes, "Sentiment analysis in Twitter," in *Proc. Semantic Eval.*, San Diego, CA, USA, Jun. 2016, pp. 215–219.

[61] L. Breiman, "Random forests," *Mach. Learn.*, vol. 45, no. 1, pp. 5–32, Oct. 2001.

[62] L. Gatti, M. Guerini, and M. Turchi, "SentiWords: Deriving a high precision and high coverage lexicon for sentiment analysis," *IEEE Trans. Affect. Comput.*, vol. 7, no. 4, pp. 409–421, Oct. 2016.

[63] S. Alghowinem *et al.*, "Multimodal depression detection: Fusion analysis of paralinguistic, head pose and eye gaze behaviors," *IEEE Trans. Affective Comput.*, vol. 1, no. 9, pp. 1–14, Dec. 2016.

[64] C. J. C. Burges, "A tutorial on support vector machines for pattern recognition," *Data Mining Knowl. Discovery*, vol. 2, no. 2, pp. 121–167, Jan. 1998.

[65] G. E. Sakr, I. H. Elhajj, and H. A.-S. Huijer, "Support vector machines to define and detect agitation transition," *IEEE Trans. Affect. Comput.*, vol. 1, no. 2, pp. 98–108, Jul./Dec. 2010.

[66] J. Marquardt *et al.*, "Age and gender identification in social media," in *Proc. Eval. Labs*, Sheffield, U.K., Sep. 2014, pp. 1129–1136.

[67] M. Koppel, S. Argamon, and A. R. Shimoni, "Automatically categorizing written texts by author gender," *Literary Linguistic Comput.*, vol. 17, no. 4, pp. 401–412, Nov. 2002.

RITA GEORGINA GUIMARÃES is currently pursuing the master's degree with the Federal University of Lavras, Brazil. She has a solid knowledge in computer science. Her current research interests include social networks, sentiment and affective analysis, web programming languages, computer networks, and recommendation systems.

RENATA L. ROSA received the M.S. degree from the University of São Paulo in 2009 and the Ph.D. degree from the Polytechnic School of the University of São Paulo, in 2015 (EPUSP). She is currently an Adjunct Professor with Department of Computer Science, Federal University of Lavras, Brazil. Her current research interests include computer networks, quality of experience of multimedia service, social networks and recommendation systems.

DENISE DE GAETANO is currently pursuing the master's degree with The University of Ulster, Coleraine, Northern Ireland. She has a solid knowledge in computer science. Her current research interests include big data, social networks, and neural networks.

DEMÓSTENES Z. RODRíGUEZ (M'12–SM'15) received the B.S. degree in electronic engineering from the Pontifical Catholic University of Peru, the M.S. degree and Ph.D. degree from the University of São Paulo in 2009 and 2013. He is currently an Adjunct Professor with the Department of Computer Science, Federal University of Lavras, Brazil. He has a solid knowledge in Telecommunication Systems and Computer Science based on 15 years of Professional experience in major companies. His research interest includes QoS and QoE in Multimedia services, Digital TV and architect solutions in Telecommunication Systems.

GRAÇA BRESSAN received the bachelor's and master's degrees from the Instituto de Matemática e Estatística and the Ph.D. degree in electronic engineering from the Polytechnic School of the University of São Paulo (EPUSP) in 1986. She is currently with the Computer Engineering Department, EPUSP, where she teaches and develops research on computer network, digital television, and performance analysis and has authored numerous articles and has oriented M.Sc. and Ph.D. degree students. She was a Professor with the Instituto de Matemática e Estatística. She was the Head of the Software Department with the Scopus Tecnologia, where she was responsible for the development of operating systems, network middleware and microcomputers firmware, software systems for tellers machines and ATM. Her current research interests includes computer networks and digital television focusing in the aspects of distributed systems, distributed middleware, QoS mechanisms, collaborative virtual environment, middleware for Digital TV, interactive digital TV, video-conferencing, modeling and performance analysis of networks, and application in distance education. She received the Award Décio Leal de Zagotis from EPUSP 2001.