

Classification of Age groups in Social network

CS4090 Project

Final Presentation

B Haranath, G Mallesh, P Santhosh kumar
Guided By: Sreenu Naik Bhukya

May 1, 2018

Outline

Introduction

Problem Statement

Literature Survey

Work Done

- Design

- Design

- Implementation

- Results and Analysis

Future Work and Conclusion

References

Introduction

- ▶ Nowadays with the constant use of the Internet users spend hours on social media platforms on the activities like
 - ▶ Browsing on e-commerce sites
 - ▶ Reading news about sports, journalism and entertainment
 - ▶ Expressing their opinions and sentiments in the form of comments
- ▶ Most of users do not provide their personnel details like age, gender etc

- ▶ There is so much concern and a great effort to analyze data from on-line social networks to predict information that may reflect different aspects of the current reality.
- ▶ Mainly age and gender plays the important roles in the predictions
- ▶ Furthermore, several machine learning algorithms such as Multilayer perceptron, Random forest were tested and Deep Convolutional Neural Network was found to exhibit higher performance.

Problem Statement

- ▶ To determine the characteristics of teenager and adults age groups, considering the writing style and both users history and profile.
- ▶ To show parameters used in this research can reach a high accuracy for determining age groups of twitter users.
- ▶ Validate the usefulness of the proposed model for classifying age groups.

Literature Survey

- ▶ Classifying the age groups methods falls into three categories.
 - ▶ Relation between the age groups and the characteristics of the writing
 - ▶ Sentiment analysis
 - ▶ Machine learning

Literature Survey

- ▶ Relation between the age groups and the characteristics of the writing.
 - ▶ "Gender classification of twitter data based on textual meta-attributes extraction" [1]
 - ▶ The age could actually alter the results of several analysis, some research have worked on trying to predict it
 - ▶ On Twitter, informing the age in profile description is not a common habit and therefore these studies would not provide reliable results

Literature Survey

► Sentiment analysis

- 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.
- ANEW- It was studied how the presence or not of gender information would affect the final result.
- 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.

Literature Survey

- ▶ Machine learning
 - ▶ Journal: Recurrent convolutional neural networks for text classification[2]
 - ▶ Recursive neural network used in some studies obtained good results in constructing sentence representations by building a tree structure.
 - ▶ This can reduce the effectiveness to capture the semantics of a whole sentence
 - ▶ The DCNN captures better the semantic of texts compared to Recursive neural network

Work Done

- ▶ Collected the Data Set through Facebook data crawler.
- ▶ Converted the dataset into .csv format (pre processing).
- ▶ Train the model with Training Data Set.
- ▶ Predicting the accuracy using test data set.

Data Extraction

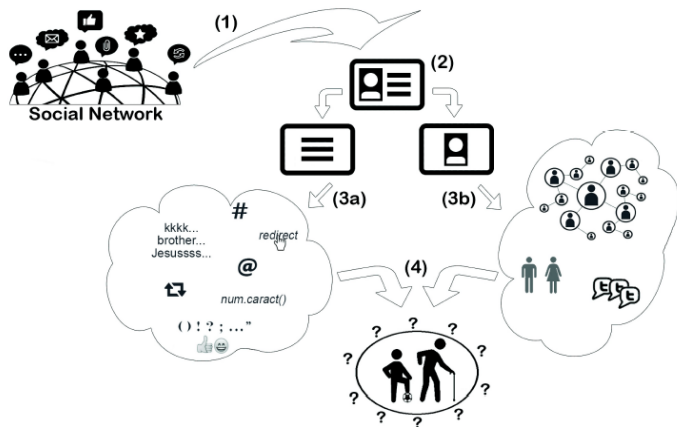


Figure: Data Extraction[3]

Extracting Features

- ▶ Message Length
- ▶ Number of Posts
- ▶ Total Post Likes
- ▶ Total comments
- ▶ Number of Friends
- ▶ Number of Page Likes
- ▶ Number of Emoji's
- ▶ Number of Hashtag's
- ▶ Label

Design

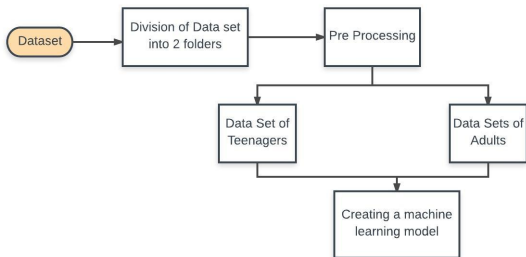


Figure: Training

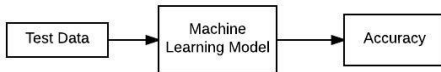


Figure: Testing

Design

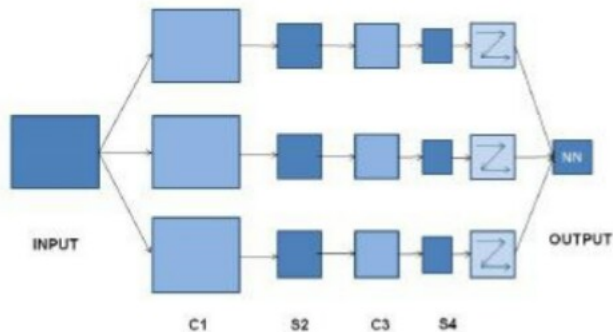


Figure: Convolutional neural network[4]

Design

- ▶ In the above figure, C1 ,C2 are Convolution Layers and S1,S2 are Max-pooling Layers.
- ▶ Proposed model for classifying the age groups have two phases
 - ▶ Data treatment extracted from social networks
 - ▶ Classification phase

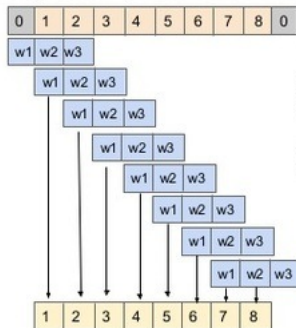
Steps in CNN Model

- ▶ STEP 1: Convolution Layer (Activation: ReLU)
- ▶ STEP 2; Max Pooling
- ▶ STEP 3: Convolution Layer (Activation: ReLU)
- ▶ STEP 4: Max Pooling
- ▶ STEP 5: Fully Connected Layer
- ▶ STEP 6: Output Layer (Activation: Sigmoid)

1-D Convolutions

1D Convolutions

When we add zero padding, we normally do so on both sides of the sequence (as in image padding)



The length result of the convolution is well known to be:
 $\text{seqlength} - \text{kwidth} + 1 = 10 - 3 + 1 = 8$

So the output matrix will be (8, 100)
because we had padding

Figure: 1-D Convolution[5]

Activation Function

- ▶ Activation Function introduce non linear properties to our Network. Their main purpose is to convert a input signal to a A-NN output signal. A Neural Network without Activation Function would simply a linear Regression Model.
- ▶ RELU:
 - ▶ It's just $R(x) = \max(0, x)$ i.e if $x \leq 0$, $R(x)=0$ and if $x > 0$, $R(x)=x$.
- ▶ Sigmoid or Logistic Activation Function:
 - ▶ It is a activation function of form $f(x) = 1 / 1 + \exp(-x)$. Its Range is between 0 and 1.
 - ▶ It is a activation function of form $f(x) = 1 / 1 + \exp(-x)$. Its Range is between 0 and 1.

Graphs of ReLu and Sigmoid Functions

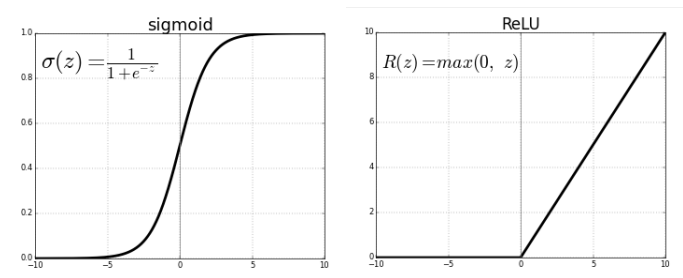


Figure: Activation functions[6]

- ▶ RELU is applied to the Hidden layers
- ▶ SIGMOID is applied to the Output Layer for Classification

Pooling

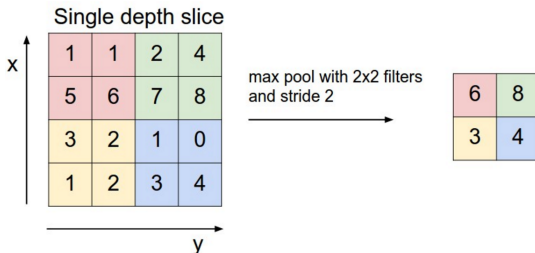


Figure: Pooling[7]

- ▶ Consider max pooling in which maximum value from a group of neurons is used in the next layer.
- ▶ Similarly average pooling takes the average value from group of neurons.

Fully Connected Layer

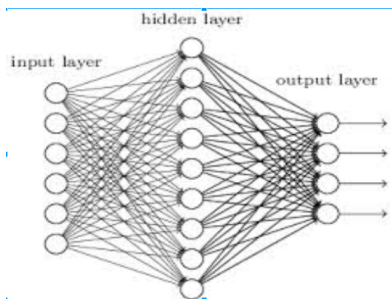


Figure: Fully Connected Layer[8]

- In a fully connected neural network, each neuron in one layer is connected to every neuron in the next layer.

Implementation

- ▶ Anaconda
 - ▶ It is a Python IDE.
 - ▶ It provides large selection of packages and commercial support.
- ▶ Tensorflow
 - ▶ TensorFlow is an machine learning library, which can be used for high-level implementation of various ML algorithms in Python.
 - ▶ However, it is used primarily for deep learning, Because Because most deep learning uses GPUs, it is common to use deep learning frameworks that implement common operations in GPUs

Implementations

- ▶ Keras

- ▶ Keras is an open source neural machine learning library written in Python.
- ▶ Deep Learning frameworks operate at 2 levels of abstraction
- ▶ **Lower Level;** This is where frameworks like Tensorflow, Theano, and PyTorch sit. This is the level where mathematical operations like Generalized Matrix-Matrix multiplication and Neural Network primitives like Convolutional operations are implemented.
- ▶ **Higher Level:** This is where frameworks like Keras sit. Generally, at this level model training are also implemented.

Parameters

PARAMETER	VALUE
Learning Rate	0.001
Number of Hidden Layers	5
Number of Neurons in a Layer	8
Epochs	50
Classifier	Sigmoid and ReLu

Figure: Parameters

Results

MODEL	ACCURACY
Deep Convolutional Neural Networks (DCNN)	94.25%
Multilayer Perceptron(MLP)	92.00%

Figure: Results

Performance

Table 4. Precision, Recall and F- measure of Proposed Model

Algorithm	Precision	Recall	F-measure
DCNN	0.9425	0.9424	0.9425
MLP	0.917	0.718	0.805

Table 5 Precision, Recall and F- measure of Reference Project

Algorithm	Precision	Recall	F-measure
DCNN	0.929	0.936	0.930
MLP	0.882	0.869	0.880

Figure: Results

Future Work and Conclusion

- ▶ At present, we are dealing with 1-D arrays. But already, Deep Learning (Convolutional Neural Networks) is dealing with images and showed an enormous progress in the field of image recognition field.
- ▶ Hence, it is more effective to implement Convnets on images rather than 1-D arrays.
- ▶ Data scientists are trying all possible chances to convert every unstructured data to structured data to achieve high progress since Convnets gives best performance for structured data.

References

- [1] J. Huang, J. Lu and C. X. Ling. Comparing Naive Bayes, Decision Trees, and SVM with AUC and Accuracy,, Proceedings of Third IEEE ICDM, 19-22 November 2003, pp. 553-556. doi:10.1109/ICDM.2003.1250975.
- [2] P. S. Vadivu and V. K. David. "Enhancing and Deriving Actionable Knowledge from Decision Trees,, IJCSIS, Vol. 8, No. 9, 2010, pp. 230-236
- [3] <https://www.semanticscholar.org/paper/Age-Groups-Classification-in-Social-Network-Using-Guimarx00E3es-Rosa/04d67280aadec8c60506e93e7d0aeef831d9e928/figure/2>
- [4] <https://www.semanticscholar.org/paper/Age-Groups-Classification-in-Social-Network-Using-Guimarx00E3es-Rosa/04d67280aadec8c60506e93e7d0aeef831d9e928/figure/0>

- [5] <https://qph.ec.quoracdn.net/main-qimg-523434af0d21bb0b59454aa9563cc90b.webp>
- [6] https://cdn-images-1.medium.com/max/1600/1*XxxiA0jJvPrHEJHD4z893g.png
- [7] <https://qph.ec.quoracdn.net/main-qimg-8afedfb2f82f279781bfefa269bc6a90.webp>
- [8] <https://www.packtpub.com/sites/default/files/Article-Images/B05478;mage001.png>