Author Identification : A Deep Approach and Comparative Study

Amitabha Mukherjee
Advisor

Ankit Pensia
12124

Anand Pandey
12109

October 15, 2015

1 Motivation

In this project, we wish to explore the problem of author identification in the field of NLP. This problem has been studied extensively using hand-designed features. We want to explore how deep learning can be used to learn the abstract and higher-level features of the document, which could be used to identify author. We wish to explore several variants from deep architecture, which could be used to tackle this problem.

2 Problem Statement

Given a text document and a set of authors, learn a function that maps the document to a single author. The training data includes the documents, labelled with their authors.

3 Previous Works

A lot of features have been suggested based on words, characters, grammar to identify authors [5][1][3].There has been a recent revival of interest in using deep learning methods for various machine learning problems and NLP[4], inorder to learn more robust features using easily available unlabelled data. Recently few architectures and models has been proposed for authorship attribution using Deep learning frameworks including LSTM, CNN, Recursive Neural Network[6][7][8]. We want to compare these architectures with traditional methods used.

4 Our Approach

We would model the sentences as a sequence of words and train different type of Neural networks to get features for a given document. This reduces the problem to multi-class

classification problem. Then we would train a multi class classifier on these feature vectors.

Among the available architectures currently present in literature, we would be using Paragraph Vector[9], Recurrent Neural Network - LSTM and Tree LSTM[2] and Recursive Neural Network.

- Although Paragraph Vector is supposed to map a sentence to a meaning space, still we would like to examine if this could act as a good feature-extractor for author attribution.
- Using RNN as architecture, and modelling sentences as a sequence of words, we can train the model using unlabelled data. While both training and inferring, we would get a feature vector for each time stamp of RNN (number of words in sentence) from the hidden layer. By using mean-pooling or weighted-averaging of these feature vectors, we will obtain a fixed-sized vector for each document and thus can train a classifier on it.

We would be comparing our results with authorship identifiers, based on stylistic hand-coded features.

5 Datasets

We propose to use different type of datsets. Since Deep learning methods require considerable amount of dataset (considerable more number of parameters), we don't expect to have good performance on PAN dataset to. We would use following datasets

- Quora Answers
- News Articles/blogs from the same domain
- PAN dataset(Training data is highly limited)(http://goo.gl/lwj1N9)

References

[1] Green, R. M., & Sheppard, J. W. Comparing Frequency- and Style-Based Features for Twitter Author Identification (2013, May). Twenty-Sixth International Florida Artificial Intelligence Research Society Conference :**64-69**

[2] K. S. Tai, R. Socher, and C. D. Manning. Improved semantic representations from tree- structured long short-term memory networks.(2015),ACM Computing Research Repository Vol: abs/1503.00075

[3] Patrick Juola. Authorship attribution. Foundations and Trends in Information Retrieval , 1(3):233334, 2006

[4] Ronan Collobert, Jason Weston, L eon Bottou, Michael Karlen, Koray Kavukcuoglu, and Pavel Kuksa. Natural language processing (almost) from scratch. The Journal of Machine Learning Research , 12:24932537, 2011.

[5] Stamatatos, E. (2009), A survey of modern authorship attribution methods. J. Am. Soc. Inf. Sci., 60: 538556. doi: 10.1002/asi.21001

[6] Pranav Jindal, Ashwin Paranjape. Deanonymizing Quora Answers (2015).CS224,Stanford University

[7] Stephen Macke, & Jason Hirshman Deep Sentence-Level Authorship Attribution (2015). CS224,Stanford University

[8] Dylan Rhodes(2015) Author Attribution with CNNs CS224,Stanford University

[9] Quoc Le, Tomas Mikolov ;Distributed Representations of Sentences and Documents arXiv:1405.4053