Text Regression on Movie Subtitles

Brian Bechtel, Jonathan Stites
University of Texas at Austin

Abstract

Below are guidlines on how to write-up your report for the final project. Of course, for a short class project, all of the comments may not be relevant. However, please use it as a general guide in structuring your final report.

A "standard" experimental NLP paper consists of the following sections:

1 Introduction

Motivate and abstractly describe the problem you are addressing and how you are addressing it. What is the problem? Why is it important? What is your basic approach? A short discussion of how it fits into related work in the area is also desirable. Summarize the basic results and conclusions that you will present.

2 Problem Definition and Algorithm

2.1 Task Definition

Precisely define the problem you are addressing (i.e. formally specify the inputs and outputs). Elaborate on why this is an interesting and important problem.

2.2 Algorithm Definition

Describe in reasonable detail the algorithm you are using to address this problem. A psuedocode description of the algorithm you are using is frequently useful. Trace through a concrete example, showing how your algorithm processes this example. The example should be complex enough to illustrate all of the important aspects of the problem but simple enough to be easily understood. If possible, an intuitively meaningful example is better than one with meaningless symbols.

3 Experimental Evaluation

3.1 Methodology

What are criteria you are using to evaluate your method? What specific hypotheses does your experiment test? Describe the experimental methodology that you used. What are the dependent and independent variables? What is the training/test data that was used, and why is it realistic or interesting? Exactly what performance data did you collect and how are you presenting and analyzing it? Comparisons to competing methods that address the same problem are particularly useful.

3.2 Results

Present the quantitative results of your experiments. Graphical data presentation such as graphs and histograms are frequently better than tables. What are the basic differences revealed in the data. Are they statistically significant?

3.3 Discussion

Is your hypothesis supported? What conclusions do the results support about the strengths and weaknesses of your method compared to other methods? How can the results be explained in terms of the underlying properties of the algorithm and/or the data.

4 Related Work

Answer the following questions for each piece of related work that addresses the same or a similar problem. What is their problem and method? How is your problem and method different? Why is your problem and method better?

5 Future Work

What are the major shortcomings of your current method? For each shortcoming, propose additions or enhancements that would help overcome it

6 Conclusion

Briefly summarize the important results and conclusions presented in the paper. What are the most important points illustrated by your work? How will your results improve future research and applications in the area?

References

[1] Mahesh Joshi, Dipanjan Das, Kevin Gimpel, Noah A. Smith, Movie Reviews and Revenues: An Experiment in Text Regression, Proceedings of NAACL-HLT, 2010.

[2] Zsolt Bitvai, Trevor Cohn, Non-Linear Text Regression with a Deep Convolutional Neural Network, Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing, 2015.

[3] Mahesh Joshi, Dipanjan Das, Kevin Gimpel, Noah A. Smith, "Movie$ Data", Web, Accessed 08 April 2017, https://www.cs.cmu.edu/ ark/movie$-data/

[4] "The Internet Movie Script Database", Web, Accessed 08 April 2017, http://www.imsdb.com/disclaimer/