

WS 2016 Project 2

Anonymous

1. INTRODUCTION

There are many examples of people extracting the sentiment of sentences using machine learning. The goal of this report is to examine how these methods translates into product specific contexts and compare the results. This is done using product reviews from brickset.com[1], using the data set found here[3] and using the python package textblob[5] and the deep learning tool nlp[4].

2. METHODOLOGY

All preprocessing has been done in Python[2]. Training and testing using textblob has also been done in python. The deep learning tool provided by Stanford nlp does most of the processing for us after the data have been preprocessed by Python. The code consist of the modules that correspond to the following sections

2.1 General pre-processing of data

The product review data provided in the google sheet[3] is missing data. The data that is not complete is filtered so only valid data, with a valid sentiment data point is used for further investigation.

This step reduces the data from 4901 data points to 3893 data points

2.2 Textblob sentiment

Before the Textblob model could be trained, the data had to go through some extra preprocessing. To speedup the training process, every product review is converted to a textblob object.

The data is then split using Numpys[?] Kfold function to prepare for 3 fold validation. The model is then trained using the three data subsets provided by the Kfold function, and then tested using the test subsets provided by the same Kfold call.

2.3 Stanford recursive deep learning for sentiment analysis

3. FINDINGS

4. CONCLUSIONS

5. REFERENCES

- [1] Brickset reviews. <http://brickset.com/reviews/eurobricks>, 2016. [Online; accessed 19-July-2008].
- [2] Python interpreter. <https://www.python.org/>, 2016. "[Online; accessed 20-mar-2016]".
- [3] Review data. <https://docs.google.com/spreadsheets/d/1Bfq2fPRiwwCGS6JgpyrVP40SyPuhuUs077xnFKWSoQ/edit?usp=sharing>, 2016. "[Online; accessed 20-mar-2016]".
- [4] Stanford nlp, deep learning sentiment analysis. <http://nlp.stanford.edu/sentiment/>, 2016. "[Online; accessed 20-mar-2016]".
- [5] Text analysis for python. <https://textblob.readthedocs.org/en/dev/>, 2016. "[Online; accessed 20-mar-2016]".

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, to republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.

Web Science 2016 DIKU, Denmark

Copyright 200X ACM X-XXXXX-XX-X/XX/XX ...\$10.00.