

Lancaster
University



My past, present and potential future of releasing software with my publications

Andrew Moore

October 3, 2017

School of Computing and Communications, Lancaster University.

SemEval publication on predicting sentiment in financial news headlines[2]

The screenshot shows the GitHub repository page for 'apmoore1 / semeval'. The repository is described as 'SemEval 2017 task 5 track 2 codebase'. It has 27 commits, 1 branch, 2 releases, 1 contributor, and is licensed under GPL-3.0. The repository contains several files and folders, including 'examples', 'final_output', 'lstms', 'models/word2vec_models', 'paper', 'presentation', 'results', and 'svrs'. The 'presentation' folder is highlighted as containing 'semeval slides'.

Repository: [apmoore1 / semeval](#) Unwatch 2 Star 2 Fork 2

Code Issues 0 Pull requests 0 Projects 0 Wiki Settings Insights

SemEval 2017 task 5 track 2 codebase. Edit

[Add topics](#)

27 commits 1 branch 2 releases 1 contributor GPL-3.0

Branch: master New pull request Create new file Upload files Find file Clone or download

File	Description	Latest commit
examples	Added all of the metrics used within SemEval.	7 months ago
final_output	First commit.	7 months ago
lstms	Added all of the metrics used within SemEval.	7 months ago
models/word2vec_models	How to use the word2vec models.	7 months ago
paper	Paper associated to the code.	6 months ago
presentation	semeval slides.	2 months ago
results	Update of install instructions.	2 months ago
svrs	Corrected the best SVR configurations and update the results accordin...	7 months ago

How people can find the code

1. Link in the paper.
2. Via the research directory

Lancaster A at SemEval-2017 Task 5: Evaluation metrics matter: predicting sentiment from financial news headlines

Contribution in Book/Report/Proceedings › Conference contribution

Published

Overview

Cite this



Andrew Moore
Paul Edward Rayson

Publication date	4/08/2017
Host publication	Proceedings of the 11th International Workshop on Semantic Evaluations (SemEval-2017)
Place of Publication	Stroudsburg, PA
Publisher	Association for Computational Linguistics
Pages	581-585
Number of pages	5
ISBN (Print)	9781945626555
Original language	English

Computing and Communications
Data Science Institute

Associated organisational unit
UCREL - University Centre for Computer
Corpus Research on Language

Electronic data

 [lancaster-semeval-2017](#) 
Accepted author manuscript, 155 KB, PDF-document

Available under license: CC BY: Creative Commons Attribution 4.0 International License

Links

 <http://aclanthology.info/papers/S17-2095/lancaster-a-at-semeval-2017-task-5-evaluation-metrics-matter-predicting-sentiment-from-financial-news-headlines>
Final published version

 <https://github.com/apmoore1/semeval>
Other version
License: GNU GPL

Text available via DOI:

 [10.18653/v1/S17-2095](https://doi.org/10.18653/v1/S17-2095)
Final published version

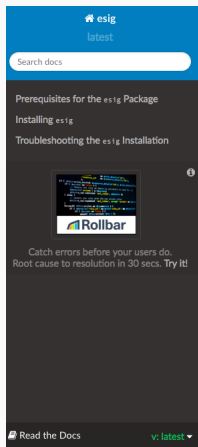
The problems that I found from my past:

1. The code does not prove what I have done has been implemented correctly.
2. The code lacks detailed documentation.
3. The code could be easier to find.

My solutions to these problems:

1. Create test e.g. unit test.
2. Create detailed documentation like readthedocs.
3. My profile on the research directory to include software tab.

esig python package¹



Docs » The esig Python Package

[Edit on GitHub](#)

The esig Python Package

This is the online documentation for the Python `esig` package. Below you'll find the table of contents.

- Prerequisites for the `esig` Package
 - Getting your Python Version
 - Installing and Configuring Boost
 - Windows
 - Linux and macOS
- Installing `esig`
 - Custom Library and Include Paths
 - Building/Installing Fails!
 - Running Tests
- Troubleshooting the `esig` Installation
 - Unknown Command `bdist_wheel`
 - Permission Denied when Installing
 - Can't Load Boost libraries
 - `numpy` error

`esig`: What is it?

The `esig` package provides implementations of a series of mathematical tools for the handling of

¹<http://esig.readthedocs.io/en/latest/>

Research directory example

Andrew Moore

Research student

Profile

Publications



Computing and Communications

InfoLab21 Lancaster University Bailrigg Lancaster

LA1 4WA

LA1 4WA

United Kingdom

E-mail: [Show email](#)

ORCID: [0000-0002-3395-0841](#)



Publications & Outputs

Domain adaptation using stock market prices to refine sentiment dictionaries



Published

Moore, A., Rayson, P. E. & Young, S. E. 23/05/2016 *Proceedings of the 10th edition of Language Resources and Evaluation Conference (LREC2016)*. European Language Resources Association (ELRA), 4 p.

Contribution in Book/Report/Proceedings > Conference contribution

Lancaster A at SemEval-2017 Task 5: Evaluation metrics matter: predicting sentiment from financial news headlines



Published

Moore, A. & Rayson, P. E. 4/08/2017 *Proceedings of the 11th International Workshop on Semantic Evaluations (SemEval-2017)*. Stroudsburg, PA: Association for Computational Linguistics, p. 581-585 5 p.

Contribution in Book/Report/Proceedings > Conference contribution

What I would like to see more of:

1. More researchers releasing their code.

The reasons I hear why researchers don't release their code:

1. I don't have the time.
2. I don't like the way the code is at the moment.
3. It is my code/ I don't want to.

1. Releasing the code no matter what it looks like is better than not.
2. When it is released others might help you.
3. Research Software Engineers (RSE)
4. More time/money

Spending more time/money is better

Reasons for RSE and more time on making the code easier to use:

1. Higher impact for your research
2. Allow other researchers to perform their job faster

Example

Stanford CoreNLP[1]² has:

1. 1711 citations
2. 3771 stars and been forked 1409 times on Github

,

²<https://stanfordnlp.github.io/CoreNLP/>

Questions?

Andrew Moore

@apmoore94

Slides: <https://github.com/apmoore1/software-as-data>

References I

- [1] C. Manning, M. Surdeanu, J. Bauer, J. Finkel, S. Bethard, and D. McClosky.
The stanford corenlp natural language processing toolkit.
In *Proceedings of 52nd Annual Meeting of the Association for Computational Linguistics: System Demonstrations*, pages 55–60. Association for Computational Linguistics, 2014.
- [2] A. Moore and P. Rayson.
Lancaster a at semeval-2017 task 5: Evaluation metrics matter: predicting sentiment from financial news headlines.
In *Proceedings of the 11th International Workshop on Semantic Evaluation (SemEval-2017)*, pages 581–585. Association for Computational Linguistics, 2017.