

Sample UCL-styled A0 scientific poster \LaTeX

Helen Wilson (with credit to Ed Long, CoMPLEX)
University College London

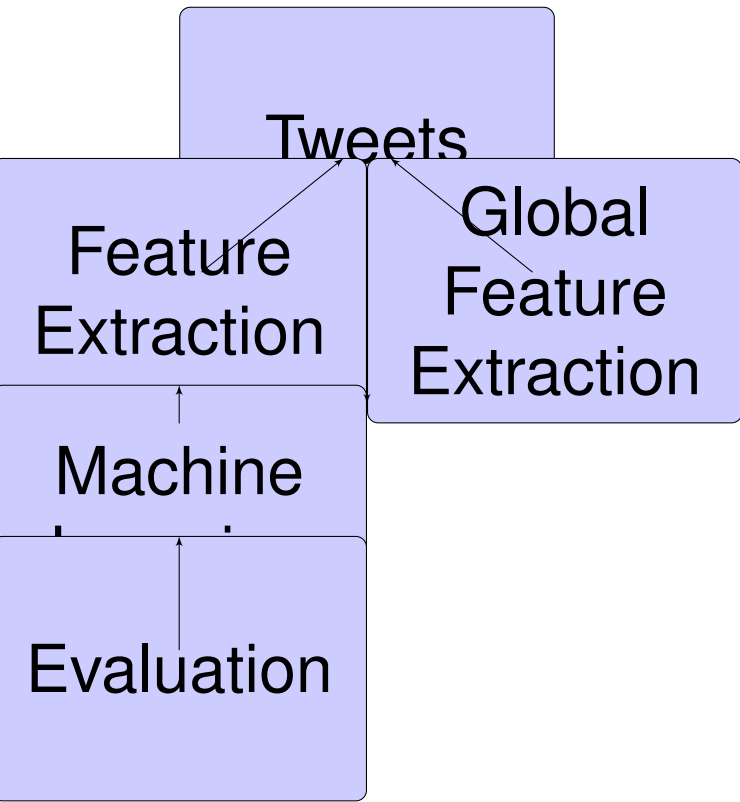


Introductory segment

Lorem ipsum dolor sit amet, consectetur adipiscing elit. Vestibulum justo. Praesent leo. Sed consectetur. Aenean pretium, diam quis mattis porttitor, elit velit scelerisque sapien, sed convallis ipsum lectus non neque. Morbi in mi eu neque luctus scelerisque. Curabitur odio. Mauris a mi. Aenean iaculis erat vel sapien. Curabitur nulla velit, feugiat quis, imperdiet sit amet, vestibulum ac, diam. Suspendisse a metus. Pellentesque vulputate venenatis eros. In auctor, eros nec sodales faucibus, erat nisl facilisis nisl, ut rutrum nisl nunc eu est. Quisque eget ante at nunc varius ultrices.

First Piece of Content

It is worth fiddling around a bit with positioning of the text blocks to get the spacing even. I like a vertical gap of about 0.4 “block units” between the horizontal bar at the end of one block and the beginning of the next. There are contruction lines at the end of the \TeX file to help with this.



Figures (and labels using the \LaTeX picture environment) work as you would expect.

Feature Extraction

The key objective of this project is to use information contain within Twitter posts to predict the mood of the markets towards certain stocks. Tweets contain a great deal of information including, the text itself, linking between users, linking to entities (either explicitly view a tag or plain text), network effect via re-tweets.

In their raw form, Tweets are not convinient for inference. As such we will focus a great deal of our time distilling the information contained within the Tweets into a form that we can use easily, ensuring all important aspects of the Tweets are preserved. This approach gives a nice layers of abstraction between the Information Retrieval and the Machine Learning steps in the pipeline. We will be gathering these features using two independent MapReduce jobs.

Global Statistics

Firstly, we will run a global job to gather statistics that depend on the entire dataset (for example tf/idf). The statistics available and the manor with which they are gathered will be a significant area of experimentation.

Some statistics that we plan to start with are:

- user average sentiment - to help us determine if someone good opinion of an entity is only because they are generally positive. This could be global or regarding a specific company.
- inverse document frequency - to be used with term frequency within tweets to potentially highlight particularly significant phrases

Use of Colour

With the `color` package you can use as much colour as you like: but the **Red** and **DarkBlue** colours defined here are more useful than the obvious **red** and **blue** versions, which tend to seem too bright when printed.

Use spacing as well as colour to highlight a key question or issue:

Who will rid me of this turbulent priest?

Praesent elementum malesuada mauris. Duis aliquet dolor ut nunc. Pellentesque euismod augue a odio.

Equations and Tables

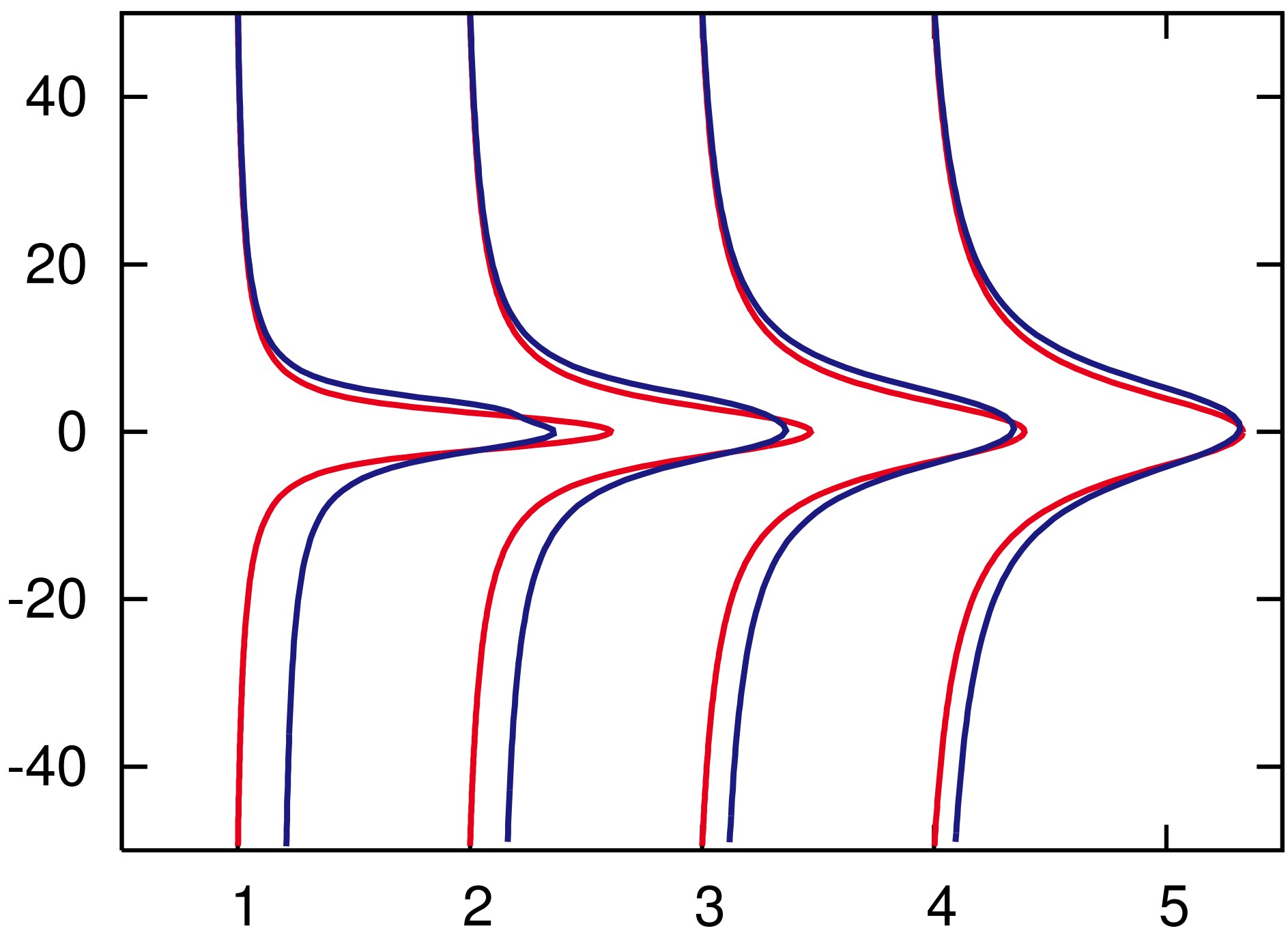
Of course, the principal reason for choosing \LaTeX is its ability with equations:

$$U_{\text{doublet}} = \frac{mg}{2\pi\mu a} \left(\int_0^\infty \left\{ 1 - \frac{2\sinh^2 s - 2s^2}{\sinh 2s + 2s} \right\} ds \right)^{-1} \approx \frac{1.55mg}{6\pi\mu a},$$

but it's equally easy to include tables:

| L | U_1 | | U_2 | | U_3 | | U_3 error |
|------|---------|---------|---------|---------|---------|---------|----------------|
| | MR | SD | MR | SD | MR | SD | |
| 2.01 | 0.65528 | 0.64739 | 0.63461 | 0.62691 | 0.00498 | 0.00451 | 9% |
| 2.10 | 0.73857 | 0.73126 | 0.59718 | 0.58784 | 0.03517 | 0.02570 | 27% |
| 2.50 | 0.87765 | 0.87482 | 0.49545 | 0.48829 | 0.07393 | 0.05853 | 21% |
| 3.00 | 0.93905 | 0.93806 | 0.41694 | 0.41356 | 0.07824 | 0.06970 | 11% |
| 4.00 | 0.97964 | 0.97945 | 0.31859 | 0.31774 | 0.06925 | 0.06639 | 4% |
| 6.00 | 0.99581 | 0.99579 | 0.21586 | 0.21575 | 0.05078 | 0.05019 | 1% |

and, of course, images:



Performance Evaluation

ML algorithm perfomance

To demonstrate the effectiveness of our learning algorithms we proceed by spiting the data into testing and training sets. Than we predict the future stock price using the training set and compute the mean squared error of the prediction using the test set.

By carrying out the above procedure with different sets of test and training data we will be able to establish the predictive power of our model as well as determine how far into the future we will be able to make effective predictions.

If the system will be used for training than the complex regression problem can be reduced to a simpler classification problem. Instead of predicting the exact price we will instead classify whether stock is going to go up or down after a specific amount of time. When evaluating this approach we intend to use the number of misclassification as a metric for determining the predictive power of our algorithm.

Back-testing

While mean square error in prediction is useful in evaluating algorithm effectiveness, low mean square error does not directly translate into trading performance.

The simple price model will need be further extended to incorporate stock liquidity and ensure that gains can be realised. Further optimisation is possible by incorporating trading fees as well as liquidity rebates to ensure that the system not only maximises prediction power but also profitability.

Discussion and Conclusions

Proin dignissim nunc in nulla. Vivamus non leo. Nulla ultrices tempor dui. Curabitur nec metus. Aliquam sed libero. Cras orci odio, molestie a, suscipit in, placerat vel, nunc. Vestibulum congue, nunc in faucibus scelerisque, ante tortor dapibus nibh, eu tristique diam urna ac magna. Proin cursus. Morbi quam ligula, fermentum vel, dapibus sit amet, euismod nec, justo. Suspendisse potenti. Nulla eu elit. Pellentesque quam est, pretium ac, suscipit sed, viverra id, sapien. Donec tempor semper tortor. Nunc vulputate. Aliquam vitae metus ut sem euismod accumsan. Duis tincidunt lacus sed ipsum. Class aptent taciti sociosqu ad litora torquent per conubia nostra, per inceptos himenaeos. Quisque nec nisl at erat ornare tempus. Etiam eros odio, ultricies non, hendrerit et, vestibulum in, felis. Vivamus gravida.

- First concluding point; we expected this to be so because of the contruction of the argument and blah.
- Second concluding point: this one is counterintuitive but we can justify it by reference to the extended discussion above.
- Third and **most important concluding point**: this is the one we're excited about.