Digital Scholarship in the Humanities, 2025, **40**, 277–279
https://doi.org/10.1093/llc/fqaf011
Advance access publication 24 February 2025
Short Paper

A note on applying the *Syuzhet* program to film dialogue

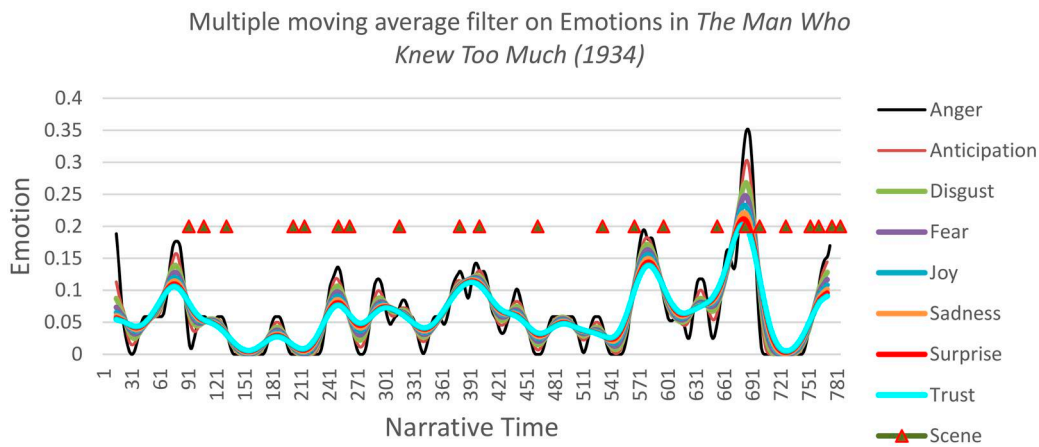
Barry Salt [1,*]

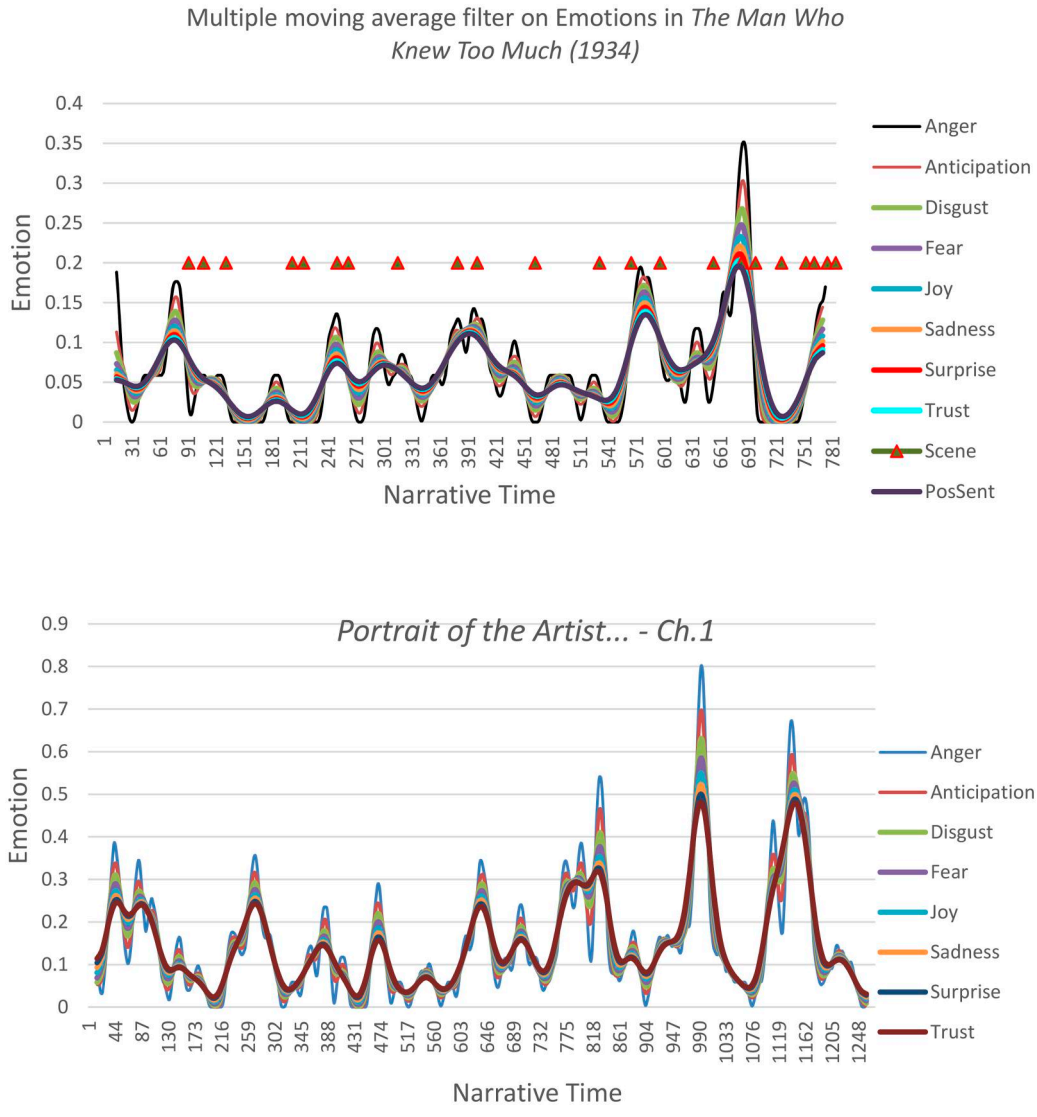
[1]Independent Scholar, London, SW13 8AH, United Kingdom
*Corresponding author. Barry Salt, Flat 35 Handel Mansions, 94 Wyatt Drive, London, SW13 8AH, United Kingdom. E-mail: barrysalt@postmaster.co.uk

Abstract

The Syuzhet program is tested in two particular cases, and produces highly questionable results.

Matthew Jockers' *Syuzhet* analysis has been cogently criticized by a number of people on theoretical grounds (see Kim 2022 for a survey and references to these criticisms), but there has been a shortage of attempts to put it to practical work in new areas. I was interested in seeing if the *Syuzhet* program (Jockers 2015) could connect emotions with the plot situations in film dramaturgy that I had investigated in pages 362 to 387 of my book *Moving Into Pictures* (Salt 2006). But the result of applying *Syuzhet* to film dialogue raises some more serious questions about the program. I took the dialogue from Alfred Hitchcock's *The Man Who Knew Too Much* (1934), and fed it into the Syuzhet program to get the values of the emotions it registers for each successive sentence. These values then had a low-pass filter applied to them. The low-pass filter I constructed for this uses a series of centred rolling averages (not the usual trailing rolling average) with widths of 5, 9, and 17 sentences applied in succession to the data. This cut-off was chosen after visual inspection of the clumping of those sentences which had some emotion, against those with none. This smoothing filter seems to produce a result similar to applying a DCT transform with a low pass size of 30. The combined plot of the eight emotions resulting is shown

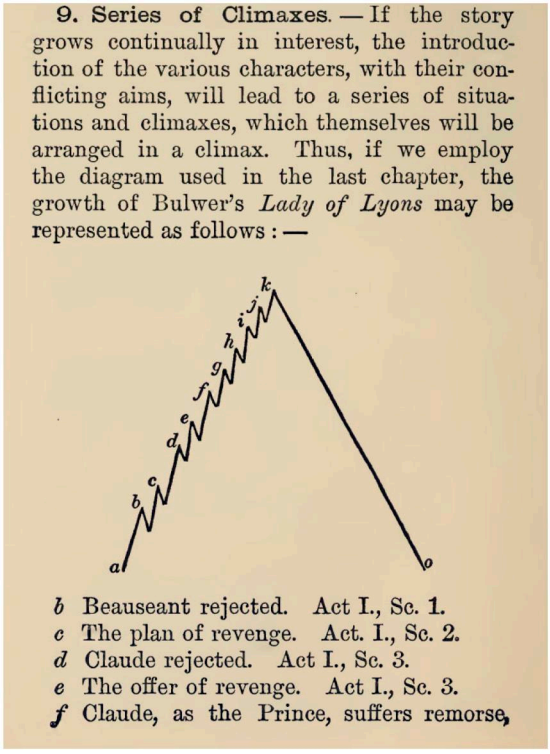




below. This plot also has the major scene divisions in the film marked on it.

The most striking thing about the graph is the near coincidence of the rise and fall of the separate emotions over the length of the film. An inspection of a table of the actual values given to each emotion by the program shows that more than a third of the sentences register no emotions at all, while on the other hand a minority of sentences show multiple emotions for each. For instance, the ninth sentence, which is "Good thing you didn't lose your head." records values for all eight emotions. On the other hand, sentence number 16 "Your fault, terrible woman!" ticks the boxes for Anger, Disgust, Fear and Sadness, which is a different emotion for each word in the sentence. The cumulative effect of hundreds of occurrences of this sort of thing down the length of the film gives the result you see in the graph. There is hardly any differentiation between the traces for the various emotions, except that the "trust" emotion misses following the smaller maxima and minima along the length of the text.

Much the highest peak on this graph is around the climax of the plot at sentence 688, which is the point at which the attempted assassination of an important



member of the audience takes place during a concert at the Albert Hall.

If we add "positive sentiment" from *Syuzhet*, it can be seen that the emotions give little new information to that available from the sentiment record.

Repeating the first stage of my process for the first chapter of James Joyce's *Portrait of the artist as a Young Man* gives the following result:

As with *The Man Who Knew Too Much*, the emotions are not very much differentiated one from another.

In general, the interesting narrative interactions of thought and deed in a story work down near sentence level, and not at the broader scale that most people have been investigating so far with *syuzhet*.

In *Portait of the Artist* … only a quarter of the text is dialogue, so I would expect a narrative text that contains no dialogue would similarly show hardly any differentiation between the emotions.

By the way, I note that there is a resemblance between these emotion graphs and the conceptual graphs of emotion in a drama that are found in playwriting manuals from the end of the nineteenth century. The example below is from page 116 of the most successful of these manuals, Alfred Hennequin's *The Art of Playwriting* (1890).

I believe the failure of the *Syuzhet* program to differentiate the emotions in my particular application is not in the program itself, but with the NRC lexicons it uses (Mohammad and Turney 2013). I expect that Tyler Rinker's *sentimentr* program will show the same problem, as it uses the same lexicons.

Author contributions

Barry Salt (Conceptualization, Data curation, Formal analysis, Investigation, Resources)

References

Jockers, M. (2015) 'Syuzhet: Extract Sentiment and Plot Arcs from Text', https://github.com/mjockers/syuzhet

Kim, H. (2022) Sentiment Analysis: Limits and Progress of the *Syuzhet* Package and Its Lexicons', *Digital Humanities Quarterly*, 16.

Mohammad, S., and Turney, P. (2013) 'Crowdsourcing a Word-Emotion Association Lexicon', *Computational Intelligence*, **29**: 436–65.

Salt, B. (2006) *Moving Into Pictures*. Starword (free download from starword.com).

Digital Scholarship in the Humanities, 2025, 40, 277–279
https://doi.org/10.1093/llc/fqaf011
Short Paper