

RESEARCH ARTICLE

# Emotional Sentence Annotation Helps Predict Fiction Genre

Spyridon Samothrakis\*, Maria Fasli

Institute for Analytics and Data Science, University of Essex, Wivenhoe Park, Colchester CO4 3SQ, United Kingdom

\* [ssamot@essex.ac.uk](mailto:ssamot@essex.ac.uk)

## Abstract

Fiction, a prime form of entertainment, has evolved into multiple *genres* which one can broadly attribute to different forms of stories. In this paper, we examine the hypothesis that works of fiction can be characterised by the emotions they portray. To investigate this hypothesis, we use the work of fictions in the Project Gutenberg and we attribute basic emotional content to each individual sentence using Ekman's model. A time-smoothed version of the emotional content for each basic emotion is used to train extremely randomized trees. We show through 10-fold Cross-Validation that the emotional content of each work of fiction can help identify each genre with significantly higher probability than random. We also show that the most important differentiator between genre novels is fear.



## OPEN ACCESS

**Citation:** Samothrakis S, Fasli M (2015) Emotional Sentence Annotation Helps Predict Fiction Genre. PLoS ONE 10(11): e0141922. doi:10.1371/journal.pone.0141922

**Editor:** Zhaohong Deng, Jiangnan University, CHINA

**Received:** April 28, 2015

**Accepted:** October 14, 2015

**Published:** November 2, 2015

**Copyright:** © 2015 Samothrakis, Fasli. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

**Data Availability Statement:** All code and data is within this repository: <https://github.com/ssamot/Novels>.

**Funding:** All funding has been provided by the Institute for Analytics and Data Science (IADS), University of Essex.

**Competing Interests:** The authors have declared that no competing interests exist.

## Introduction

Fiction forms the basis of most modern forms of entertainment; works of fiction are often “translated” into games and movies, providing a formidable substrate of stories on which one can share unique experiences. One of the most important elements of a story is the rhetoric [1] it uses to elicit different levels of emotion to its readers. Emotion seems to be fundamental to human behaviour and studies have shown that damage in emotionally-charged neural pathways causes severe issues in decision making [2]. Emotional responses to fiction go beyond the simple hedonistic outlook prevalent in modern life; this allows the reader to immerse herself in a fictional environment and potentially (re-)live and (re-)experience a life that is not her own.

Textual analysis of emotion has mostly focused on sentiment analysis, which attributes valence (i.e., positive or negative feelings) to text snippets [3]. Alongside this arguably one-dimensional view of human emotions, models have been developed that try to capture all six basic emotions (i.e., ANGER, DISGUST, FEAR, JOY, SADNESS, SURPRISE) [4] in text. In the process, a popular database of emotionally-charged (e.g. “hate”) synonyms (“synsets”) has been identified and labelled WORDNET-AFFECT [5]. Synsets are synonym rings that form WORDNET [6], a large online thesaurus. Previous work in analysing the emotions prevalent in text has focused on analysing news headlines and blog posts [7, 8] and books in time [9]. Blog texts in particular have been used [10] to help identify the mood of authors when writing each post. Though classification results over the baseline (over 0.5 precision) are not spectacular, such methods are

clearly viable. In the classic task of movie review sentiment analysis [11] for example) accuracy rates of over 0.83 are achievable.

As a written work of fiction (essentially a long piece of text) tries to convey emotions, it is natural to assume that it can be analysed using text analysis techniques and also postulate that different types of fiction portray different emotions, both in terms of time (i.e. which portion of a work of fiction portrays what) and in terms of type (what emotions are portrayed in what works of fiction). This forms our main hypothesis and the basis of this study. In this study, we identify the emotional content of sentences, move from sentences to novels and try to predict the genre of a work of fiction based on this analysis of its emotional content. To achieve this we train extremely randomized trees and compare with baseline classifiers. Results show a huge boost in predictive power compared to both stratified and most frequent class baseline classifiers.

Though (to the best of our knowledge) this is the first study trying to link emotions portrayed in fiction with the fiction genre, there are some peripheral studies worth mentioning. Bentley et. al. [12] followed an approach similar to the one used here, but on a huge historical corpus provided by google online and showed a strong correlation between economic performance and the use of negatively charged emotions. Hughes [13] et.al. use texts available in project Gutenberg to show that within certain time periods literary styles happen to coincide. In the realm of sentiment analysis, one can find similar work of a descriptive nature being done by Jockers [14], but no direct link to genre is given, albeit it should be straightforward to extend it. Within the context of Songs, blogs and speeches, Dodds and Danforth [15] are able to correlate author age and time at the time of writing with the happiness level of each piece of work. Finally, the closest piece of work we could identify to ours [16] is a quantitative and visualisation study on the role of emotions in fiction, but with no correlation with genre.

## Materials and Methods

Our basic hypothesis is that different genres of fiction elicit different emotions, and thus one can predict the genre of a work of fiction through its emotional content. In this section we describe how and what was collected alongside how we did the processing.

### Materials

We used Project Gutenberg [17] as a means to collect freely available works of fiction, categorised by individual experts. The genres, numbers and ids can be seen in Table 1. We performed a simple keyword search for each category and collected 3403 works in total. The breakdown for each genre can be seen in Table 1.

The materials were collected and cleaned up by removing copyright and Project Gutenberg notices. We went through the whole of project Gutenberg catalogue and searched for the following keywords under the **subject** metadata: “SCIENCE FICTION”, “HORROR”, “WESTERN”, “FANTASY”, “CRIME FICTION”, “MYSTERY”, “HUMOR”, “ROMANCE”. Romance novels were discarded as

**Table 1. List of genres, numbers of instances found and unique ids for each class (i.e, genre).**

MYSTERY	648	0
HUMOR	662	1
FANTASY	346	2
HORROR	108	3
SCIENCE FICTION	1252	4
WESTERN	387	5

doi:10.1371/journal.pone.0141922.t001

very few were found (< 20). We also removed all copies of the humoristic journal LONDON CHARIVARI.

## Emotional Content of fiction

We measure the emotional content of a sentence using a simple approach termed WN-AFFECT PRESENCE [8], which simply adds up the number of Synsets (i.e. rings of synonyms) in each sentence that are attributed to an emotion as provided by WORDNET-AFFECT. We use the Natural Language Toolkit in Python [18] to extract the relevant information for each sentence. Thus each sentence will get six real numbers, one for each basic emotion. Note that the presence of words is the only thing we take into consideration when calculating the emotional content of a sentence, which is arguably the simplest method possible. It obviously does not take into account many of the intricacies of text, but it seems to work relatively well in practice, performing reasonably well as has been demonstrated in previous studies [8]. WORDNET-AFFECT had precision scores of 0.33, 0, 100.0, 50, 33, 13.4 for ANGER, DISGUST, FEAR, JOY, SADNESS, SURPRISE. Recall rates were 12.8, -1.59, 24.86, 10.32, 8.56, 3.06 respectively. In the last four emotions, these precision scores ranked top among four different methods in the task of emotional news headline classification. The algorithm had low recall and performed badly on DISGUST (though this could have been due to a data bias). If we are to take into account that no overall “best” method exists, choosing a simple method that performs reasonably well as a “feature extractor” seems reasonable.

We proceed by repeating the process for each sentence of each work of fiction. Thus initially, each work is transformed into six separate signals, one for each basic emotion. The signals change through time as the emotional content of the sentences of fiction changes. Each signal is smoothed using a Hanning smoother [19] with window length being the total length of the signal divided by three,  $M = |tss|/3$ , where  $tss$  is the signal vector for the whole text. Three was chosen heuristically, however the important fact to note is that the smoothing is proportional to the total signal window, as signals have different sizes. The window smoother thus is as in Eq 1

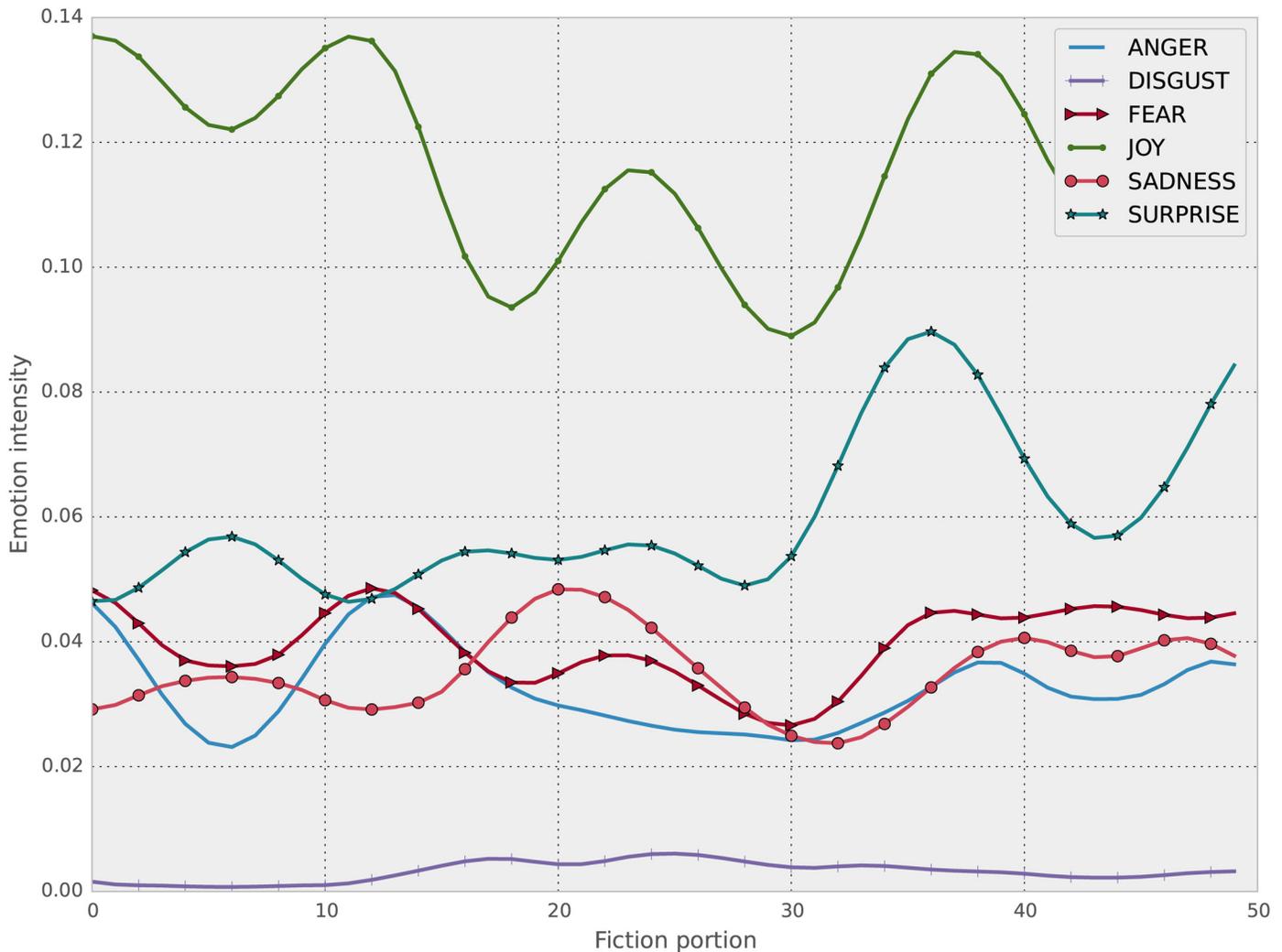
$$w(n) = 0.5 - 0.5\cos((2\pi n)/(M - 1)) \tag{1}$$

We then take the average of the signal for each  $n$  sentences, where  $|tss|/c$ , where  $c = 50$  is a constant. This in effect creates a smoothed version of the signals with 50 timesteps, no matter how big the original signal length (i.e. the size of the work of fiction) was. This however meant that some works that had less than 50 sentences in total had to be removed, which brought down the total sample size to 3377. Sample signals of this type for all emotions can be seen for two novels in Figs 1 and 2. Notice the continuous fluctuation of signal strength. The smoothing/averaging process described has the explicit goal of turning a very noisy signal to a version that can be fed into a classifier and that minor differences are removed. The noise comes from multiple sources (e.g., errors in the emotional content analysis) but the size of the overall text allows for the total feeling to be captured.

In total, we analysed 3377 individual works of fiction and we extracted 300 features for each work of fiction, 50 for each basic emotion.

## Results

We perform 10-fold cross-validation using all samples, as this was shown empirically to have an optimal bias-variance tradeoff [20]. We train extremely random forests [21] using scikit-learn’s Python implementation [22] of the algorithm. We use 1500 trees. A scaled average version (between [0, 1]) of all the confusion matrices of the results can be seen in Fig 3 and the



**Fig 1. Emotional content for MURDER AT BRIDGE, by ANNE AUSTIN, of class MYSTERY.**

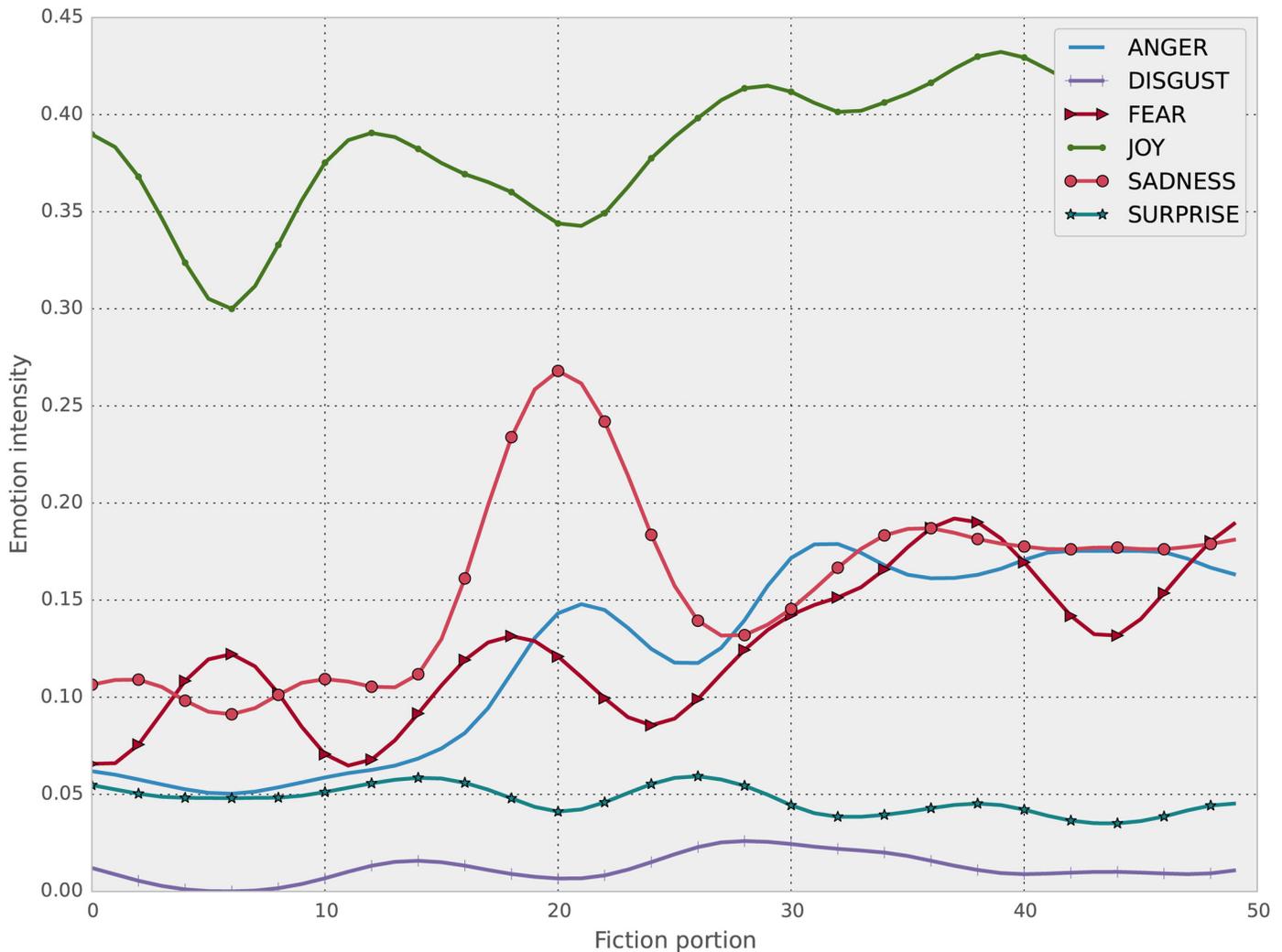
doi:10.1371/journal.pone.0141922.g001

results are available in Table 2. The stratified classifier simply chooses labels proportionally to the instances represented in the data. It creates a multinomial distribution based on the training labels and samples from it in order to predict the correct label. The most frequent classifier predicts the class with the highest number of instances. One can clearly see that using emotions can help almost double accuracy performance. In our case, accuracy is measured as:

$$accuracy(genre, predicted\_genre) = 1/3377 \sum_0^{3376} \mathbb{1}\{genre = predicted\_genre\},$$

where  $\mathbb{1}$  is the indicator function.

As evident in the confusion matrix, the classifier mostly misclassifies HORROR fiction as either FANTASY or SCIENCE FICTION. The biggest proportion of correctly classified fiction is SCIENCE FICTION, followed up by HUMOR, which also make up the bulk of our samples. We calculate the importance of each feature and present it in Fig 4. Each tree used by the classifier attributes its own split importance to features and the average of this split importance is plotted alongside the 95% confidence interval. We can see that the most discriminating features is FEAR.

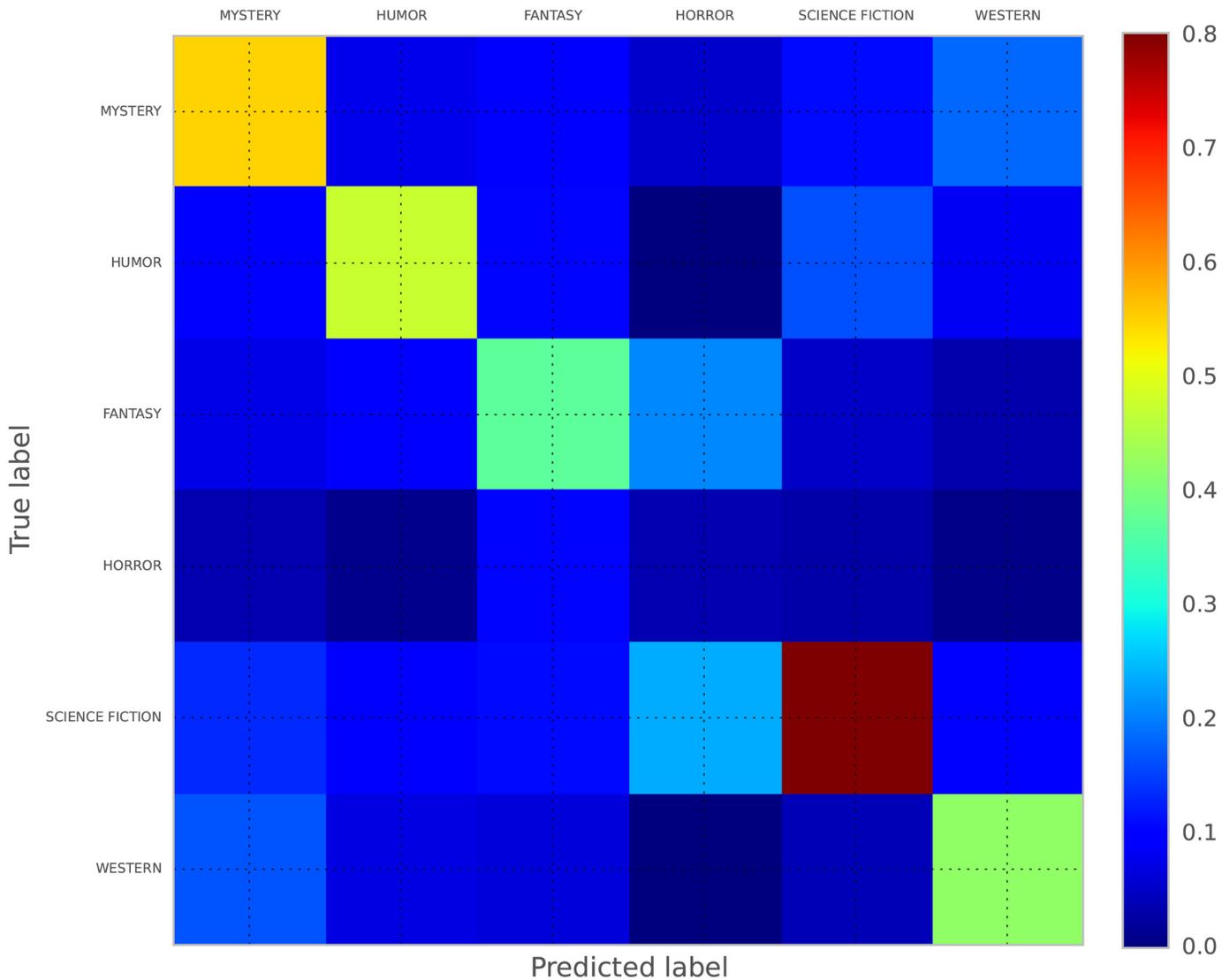


**Fig 2. Emotional content for FRANKENSTEIN, by MARY W. SHELLEY, of class HORROR.**

doi:10.1371/journal.pone.0141922.g002

It is interesting to observe that averages (as portrayed in Fig 5) cannot be correlated with genres, although some averages make intuitive sense (e.g., high FEAR and DISGUST values in HORROR fiction). Overall it seems that horror novels have higher “emotional content”, with the FEAR element spiking in the last part of the novels. On the other hand, SCIENCE FICTION novels seem comparatively low on emotions. HUMOR seems to be based heavily on SUPRISE keywords. Another interesting result that clearly stands out is how strong the JOY element is in all texts, relatively to other emotions. This might be due to the higher number of keywords related to joy compared to the rest of the emotions. Overall however, the “time element” of emotions needs further clarification, and possibly a study on more novels, in order to be able to draw more representative conclusions.

The results we present are simple and portray a clear relationship between emotions and genre. This relationship however between the emotional content and the genre is somewhat lost within the complex non-linearities of the trees used. To amend this we summarised the whole text with just a single variable, and though there is a drop in performance using trees (accuracy is  $\approx 0.54$ ), the strength is still significant. We can now fit a linear function



**Fig 3. Scaled confusion matrix.**

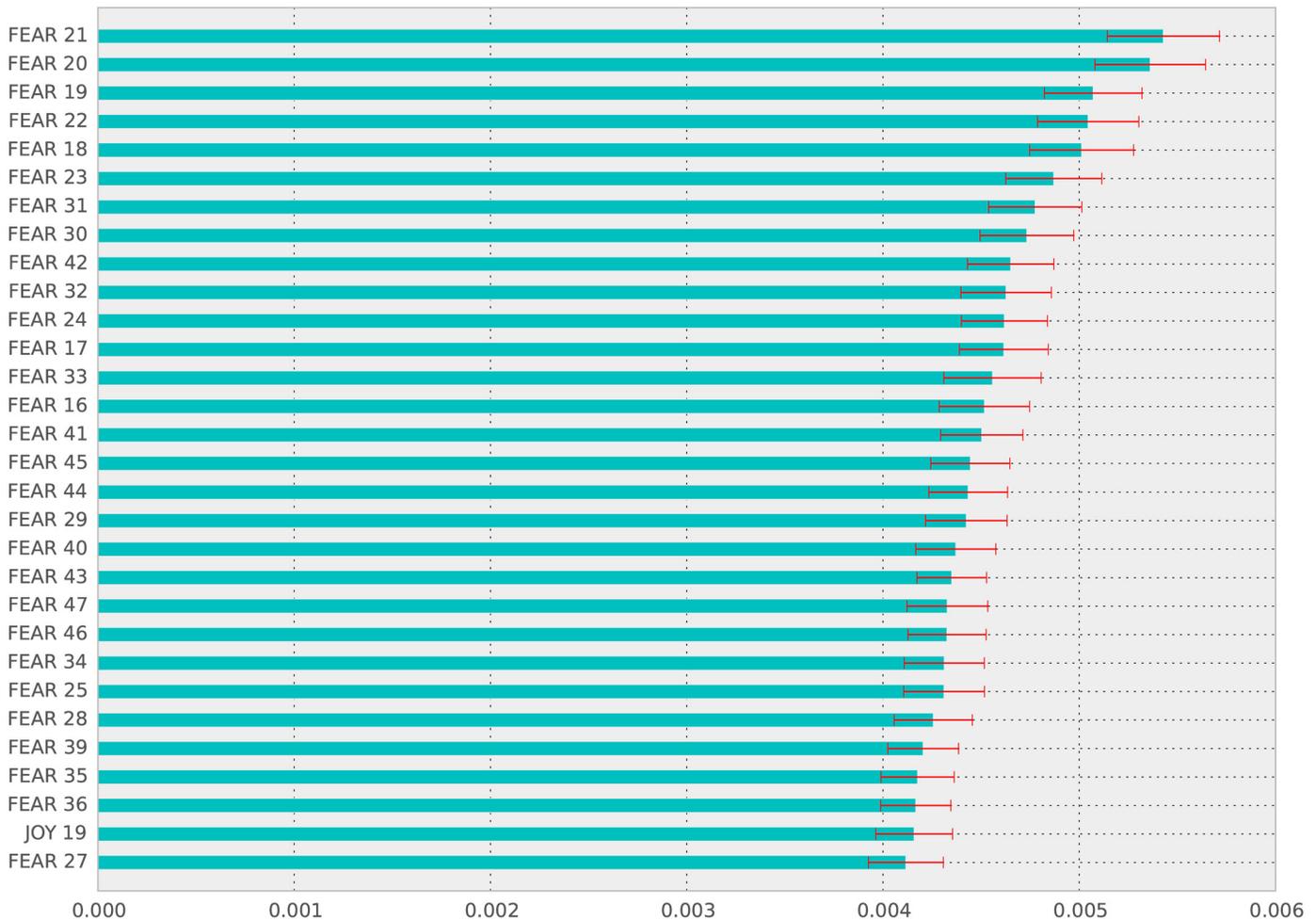
doi:10.1371/journal.pone.0141922.g003

approximator that can be analysed intuitively. Accepting a further drop in performance (this time accuracy is  $\approx 0.42$ ), using logistic regression in conjunction with “l1” regularisation, we can obtain coefficients for each emotion (see Table 3). From Table 3 we can draw some interesting conclusions; HUMOR fiction is negatively correlated with FEAR, but has a positive correlation with SURPRISE, while horror and fantasy have a strong positive correlation. On the other hand, SCIENCE FICTION has a negative correlation with JOY. This analysis further demonstrates FEAR as a discriminatory variable for works of fiction.

**Table 2. Baseline classifiers and extremely random forests accuracy.** Baseline classifiers performance is on the whole training set.

Most Frequent	0.369558779982
Stratified	0.223275096239
Extremely Random Forests	0.579646017699

doi:10.1371/journal.pone.0141922.t002



**Fig 4. Feature importances, with  $\alpha = 0.05$  confidence intervals bars.**

doi:10.1371/journal.pone.0141922.g004

## Discussion

This research was motivated by the hypothesis that works of fiction can be characterised by their emotional content and emotions can therefore be used to predict genre. We have deployed and presented the results of a simple method for correlating the emotions portrayed in works of fiction with the genre of the work. Our experiments using the Project Gutenberg works demonstrate that the use of emotions can indeed be used to predict genre with significant accuracy. As far as we are aware, this analysis of fiction texts based on emotions has not been undertaken before. We have also shown that the most important emotion, when it comes to classifying texts, is FEAR.

There is a number of important future directions one can take as part of this research. The most important one is analyse more fiction sources and vastly increase the size of the data. This would allow for a much more comprehensive study to take place and would extend the results of this limited-in-size study. Coupled with vast amounts of works of fiction, one can possibly try to analyse how genres correlate with the time they were published and the relationship of emotions to sales and user ratings, linking this work to previous studies [13].

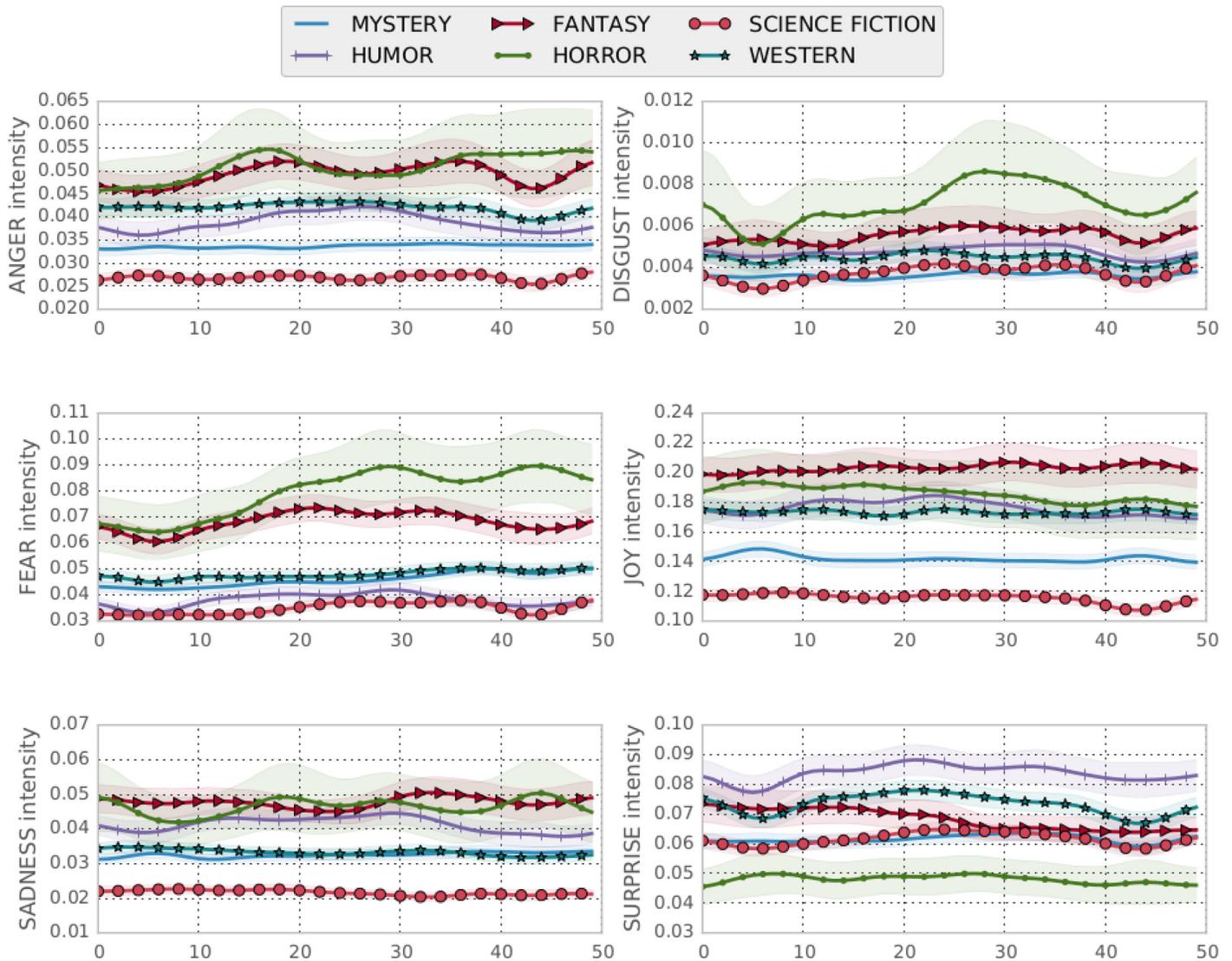


Fig 5. Average strength of each emotion among all texts. Statistical significance for  $\alpha = 0.05$ .

doi:10.1371/journal.pone.0141922.g005

Another important follow-up of this research is treating the hyperparameters used in this study (i.e, the smoothing/averaging process and their configuration) as variables and conduct an extensive study as to the impact of these variables in the quality of the learning.

Table 3. Emotion importance matrix.

	ANGER	DISGUST	FEAR	JOY	SADNESS	SURPRISE
MYSTERY	-0.066	-0.093	0.152	-0.145	0.114	-0.052
HUMOR	0.069	0.014	-0.56	0.282	0.248	0.432
FANTASY	-0.003	0	0.261	0.051	0.158	-0.082
HORROR	0.017	0.064	0.27	-0.094	0.055	-0.182
SCIENCE FICTION	-0.114	0.082	0	-0.222	-0.472	0.04
WESTERN	0.179	-0.015	-0.082	0.145	-0.15	-0.007

doi:10.1371/journal.pone.0141922.t003

Another possible avenue for exploration would be to increase the number of emotion keywords used (and thus the link between text and each book), possibly by introducing the dataset used by Saif [16]. Finally, a completely different direction would be to use an n-gram model (e.g., in the form of recurrent neural networks) to help identify genres.

All the above extensions however require access to vast quantities of digitised works of fiction and proper annotation of their genre, something that might not be publicly available for the foreseeable future. What we have done in this study is establish a new problem and create a baseline that is open up to further study in an extremely interesting topic.

## Author Contributions

Conceived and designed the experiments: SS MF. Performed the experiments: SS. Analyzed the data: SS. Contributed reagents/materials/analysis tools: SS. Wrote the paper: SS MF.

## References

1. Booth WC. The rhetoric of fiction. University of Chicago Press; 1983.
2. Damasio AR. The feeling of what happens: Body and emotion in the making of consciousness. Random House; 2000.
3. Agarwal A, Xie B, Vovsha I, Rambow O, Passonneau R. Sentiment analysis of twitter data. In: Proceedings of the Workshop on Languages in Social Media. Association for Computational Linguistics; 2011. p. 30–38.
4. Ekman P, Friesen WV. Constants across cultures in the face and emotion. *Journal of personality and social psychology*. 1971; 17(2):124. doi: [10.1037/h0030377](https://doi.org/10.1037/h0030377) PMID: [5542557](https://pubmed.ncbi.nlm.nih.gov/5542557/)
5. Strapparava C, Valitutti A, et al. WordNet Affect: an Affective Extension of WordNet. In: LREC. vol. 4; 2004. p. 1083–1086.
6. Fellbaum C. WordNet. Wiley Online Library; 1998.
7. Strapparava C, Mihalcea R. Semeval-2007 task 14: Affective text. In: Proceedings of the 4th International Workshop on Semantic Evaluations. Association for Computational Linguistics; 2007. p. 70–74.
8. Strapparava C, Mihalcea R. Learning to identify emotions in text. In: Proceedings of the 2008 ACM symposium on Applied computing. ACM; 2008. p. 1556–1560.
9. Acerbi A, Lampos V, Garnett P, Bentley RA. The expression of emotions in 20th century books. *PloS one*. 2013; 8(3):e59030. doi: [10.1371/journal.pone.0059030](https://doi.org/10.1371/journal.pone.0059030) PMID: [23527080](https://pubmed.ncbi.nlm.nih.gov/23527080/)
10. Mishne G, et al. Experiments with mood classification in blog posts. In: Proceedings of ACM SIGIR 2005 workshop on stylistic analysis of text for information access. vol. 19. Citeseer; 2005.
11. Pang B, Lee L, Vaithyanathan S. Thumbs up?: sentiment classification using machine learning techniques. In: Proceedings of the ACL-02 conference on Empirical methods in natural language processing—Volume 10. Association for Computational Linguistics; 2002. p. 79–86.
12. Bentley RA, Acerbi A, Ormerod P, Lampos V. Books Average Previous Decade of Economic Misery. *PLoS ONE*. 2014 01; 9(1):e83147. doi: [10.1371/journal.pone.0083147](https://doi.org/10.1371/journal.pone.0083147) PMID: [24416159](https://pubmed.ncbi.nlm.nih.gov/24416159/)
13. Hughes JM, Foti NJ, Krakauer DC, Rockmore DN. Quantitative patterns of stylistic influence in the evolution of literature. *Proceedings of the National Academy of Sciences*. 2012; 109(20):7682–7686. doi: [10.1073/pnas.1115407109](https://doi.org/10.1073/pnas.1115407109)
14. Jockers M. Syuzhet. GitHub; 2015. <https://github.com/mjockers/syuzhet>.
15. Dodds PS, Danforth CM. Measuring the happiness of large-scale written expression: Songs, blogs, and presidents. *Journal of Happiness Studies*. 2010; 11(4):441–456. doi: [10.1007/s10902-009-9150-9](https://doi.org/10.1007/s10902-009-9150-9)
16. Mohammad S. From once upon a time to happily ever after: Tracking emotions in novels and fairy tales. In: Proceedings of the 5th ACL-HLT Workshop on Language Technology for Cultural Heritage, Social Sciences, and Humanities. Association for Computational Linguistics; 2011. p. 105–114.
17. Hart M. Project gutenber. Project Gutenberg; 1971.
18. Bird S, Klein E, Loper E. Natural language processing with Python. “O’Reilly Media, Inc.”; 2009.
19. Tukey JW. Exploratory data analysis. Reading, Ma. 1977; 231:32.
20. Kohavi R. A Study of Cross-validation and Bootstrap for Accuracy Estimation and Model Selection. In: Proceedings of the 14th International Joint Conference on Artificial Intelligence—Volume 2. IJCAI’95. San Francisco, CA, USA: Morgan Kaufmann Publishers Inc.; 1995. p. 1137–1143.

21. Geurts P, Ernst D, Wehenkel L. Extremely randomized trees. *Machine learning*. 2006; 63(1):3–42. doi: [10.1007/s10994-006-6226-1](https://doi.org/10.1007/s10994-006-6226-1)
22. Pedregosa F, Varoquaux G, Gramfort A, Michel V, Thirion B, Grisel O, et al. Scikit-learn: Machine learning in Python. *The Journal of Machine Learning Research*. 2011; 12:2825–2830.