Letting Emotions Flow: Success Prediction by Modeling the Flow of Emotions in Books

Suraj Maharjan* Sudipta Kar* Manuel Montes-y-Gómez† Fabio A. González‡ Thamar Solorio*

*Department of Computer Science, University of Houston
†Instituto Nacional de Astrofísica Optica y Electronica, Puebla, Mexico
‡Systems and Computer Engineering Department, Universidad Nacional de Colombia

{smaharjan2, skar3}@uh.edu, solorio@cs.uh.edu
mmontesg@ccc.inoep.mx
fagonzalez@unal.edu.co

Abstract

Books have the power to make us feel happiness, sadness, pain, surprise, or sorrow. An author’s dexterity in the use of these emotions captivates readers and makes it difficult for them to put the book down. In this paper, we model the flow of emotions over a book using recurrent neural networks and quantify its usefulness in predicting the book’s success. We obtained the best weighted F1-score of 0.690 for predicting books’ success in a multitask setting (simultaneously predicting success and genre of books).

1 Introduction

Books have the power to evoke a multitude of emotions in their readers. They can make readers laugh at a comic scene, cry at a tragic scene and even feel pity or hate for the characters. Specific patterns of emotion flow within books can compel the reader to finish the book, and possibly pursue similar books in the future. Like a musical arrangement, the right emotional rhythm can arouse readers, but even a slight variation in the composition might turn them away.

Vonnegut (1981) discussed the potential of plotting emotions in stories on the “Beginning-End” and the “Ill Fortune-Great Fortune” axes. Reagan et al. (2016) used mathematical tools like Singular Value Decomposition, agglomerative clustering, and Self Organizing Maps (Kohonen et al., 2001) to generate basic shapes of stories. They found that stories are dominated by six different shapes. They even correlated these different shapes to the success of books. Mohammad (2011) visualized emotion densities across books of different genres. He found that the progression of emotions varies with the genre. For example, there is a stronger progression into darkness in horror stories than in comedy.

As an example, in Figure 1, we draw the flow of emotions across the book: Alice in Wonderland. The plot shows continuous change in trust, fear, and sadness, which relates to the main character’s getting into and out of trouble. These patterns present the emotional arcs of the story. Even though they do not reveal the actual plot, they indicate major events happening in the story.

In this paper, we hypothesize that readers enjoy emotional rhythm and thus modeling emotion flows will help predicting a book’s potential success. In addition, we show that using the entire content of the book yields better results. Considering only a fragment, as done in earlier work that focuses mainly on style (Maharjan et al., 2017; Ashok et al., 2013), disregards important emotional changes. Similar to Maharjan et al. (2017), we also find that adding genre as an auxiliary task improves success prediction.

2 Methodology

We extract emotion vectors from different chunks of a book and feed them to a recurrent neural network (RNN) to model the sequential flow of emotions. We aggregate the encoded

Figure 1: Flow of emotions in Alice in Wonderland.
GRUs will read the sequence from
marize the contextual emotion flow information
rent Units (GRUs) (Bahdanau et al., 2014) to sum-
Model
dispersion of emotions within a chunk.

The motivation behind using
chunk, respectively. The chunk vectors have a di-
mean of the sentence emotion vectors for the

where, \( n \) is the total number of sentences, \( s_{ij} \) and \( \bar{s}_i \) are the jth sentence emotion vector and the mean of the sentence emotion vectors for the ith chunk, respectively. The chunk vectors have a dimension of 20 each. The motivation behind using the standard deviation as a feature is to capture the dispersion of emotions within a chunk.

Model: We then use a bidirectional Gated Recur-
cents into a chunk emotion vector by taking the av-
Finally, we aggregate these sentence emotion vec-
from \( x_n \) to \( x_1 \), respectively. These operations will
and backward hidden states \( \overrightarrow{h}_i \) and \( \overleftarrow{h}_i \). The an-
notation for each chunk \( x_i \) is obtained by concate-
ating its forward hidden states \( \overrightarrow{h}_i \) and its back-
ward hidden states \( \overleftarrow{h}_i \), i.e. \( h_i = [\overrightarrow{h}_i; \overleftarrow{h}_i] \). We then
learn the relative importance of these hidden states for the classification task. We combine them by taking the weighted sum of all \( h_i \) and represent the final book vector \( r \) using the following equation:

\[
\begin{align*}
\alpha_i &= \frac{\exp(score(h_i))}{\sum_v \exp(score(h_v))} \\
ran = \sum_i \alpha_i h_i \\
score(h_i) &= v^T \text{selu}(W_i h_i + b_i)
\end{align*}
\]

where, \( \alpha_i \) are the weights, \( W_i \) is the weight matrix, \( b_i \) is the bias, \( v \) is the weight vector, and \( \text{selu} \) (Klambauer et al., 2017) is the nonlinear activation function. Finally, we apply linear transformation which maps the book vector \( r \) to the number of classes. In case of the single task (ST) setting, where we only predict success, we apply sigmoid activation to get the final prediction probabilities and compute errors using binary cross entropy loss. Similarly, in multitask (MT) setting, where we predict both the success and genre (Mahajan et al., 2017), we apply softmax activation to

Figure 2: Multitask Emotion Flow Model.

**Representation of Emotions:** NRC Emotion Lexicons provide ~14K words (Version 0.92) and their binary associations with eight types of elementary emotions (anger, anticipation, joy, trust, disgust, sadness, surprise, and fear) from the Hourglass of emotions model with polarity (positive and negative) (Mohammad and Turney, 2013, 2010). These lexicons have been shown to be effective in tracking emotions in literary texts (Mohammad, 2011).

**Inputs:** Let \( X \) be a collection of books, where each book \( x \in X \) is represented by a sequence of \( n \) chunk emotion vectors, \( x = (x_1, x_2, \ldots, x_n) \), where \( x_i \) is the aggregated emotion vector for the chunk \( i \), as shown in Figure 2. We divide the book into \( n \) different chunks based on the number of sentences. We then create an emotion vector for each sentence by counting the presence of words of the sentence in each of the ten different type of emotions of the NRC Emotion Lexicons. Thus, the sentence emotion vector has a dimension of 10. Finally, we aggregate these sentence emotion vectors into a chunk emotion vector by taking the average and standard deviation of sentence vectors in the chunk. Mathematically, the \( i \)th chunk emotion vector \( (x_i) \) is defined as follows:

\[
x_i = \left[ \frac{\sum_j N s_{ij}}{N}, \sqrt{\frac{\sum_j (s_{ij} - \bar{s}_i)^2}{N}} \right]
\]

where, \( N \) is the total number of sentences, \( s_{ij} \) and \( \bar{s}_i \) are the jth sentence emotion vector and the mean of the sentence emotion vectors for the \( i \)th chunk, respectively. The chunk vectors have a dimension of 20 each. The motivation behind using the standard deviation as a feature is to capture the dispersion of emotions within a chunk.

**Model:** We then use a bidirectional Gated Recurrent Units (GRUs) (Bahdanau et al., 2014) to summarize the contextual emotion flow information from both directions. The forward and backward GRUs will read the sequence from \( x_1 \) to \( x_n \), and
get the final prediction probabilities for the genre prediction. Here, we add the losses from both tasks, i.e. \( L_{\text{total}} = L_{\text{success}} + L_{\text{genre}} \) (\( L_{\text{success}} \) and \( L_{\text{genre}} \) are success and genre tasks’ losses, respectively), and then train the network using backpropagation.

3 Experiments

3.1 Dataset

We experimented with the dataset introduced by Maharjan et al. (2017). The dataset consists of 1,003 books from eight different genres collected from Project Gutenberg\(^1\). The authors considered only those books that were at least reviewed by ten reviewers. They categorized these books into two classes, Successful (654 books) and Unsuccessful (349 books), based on the average rating for the book in Goodreads\(^2\) website.

3.2 Baselines

We compare our proposed methods with the following baselines:

**Majority Class**: The majority class baseline predicts success class and obtains a weighted F1-score of 0.506 for all the test instances.

**SentiWordNet+SVM**: Maharjan et al. (2017) used SentiWordNet (Baccianella et al., 2010) to compute the sentiment features along with count of different Part of Speech (PoS) tags for every 50 consecutive sentences and used SVM classifier.

**NRC+SVM**: We concatenate the chunk emotion vectors \( (x_t) \) created using the NRC lexicons and feed them to the SVM classifier.

These baseline methods do not incorporate the sequential flow of emotions across the book and treat each feature independently of each other.

3.3 Experimental Setup

We experimented with the same random stratified splits of 70:30 training to test ratio as used by Maharjan et al. (2017). We use the SVM algorithm for the baselines and RNN for our proposed emotion flow method. We tuned the C hyperparameter of the SVM classifier by performing grid search on the values \( \{1e^{-4}, \ldots, 1e^{-1}\} \), with three fold cross validation method on the training split. For the experiments with RNN, we first took a random stratified split of 20% from the training data as validation set. We then tuned the RNN hyperparameters by running 20 different experiments with a random selection of different values for the hyperparameters. We tuned the weight initialization (Glorot Uniform (Glorot and Bengio, 2010), Le-Cun Uniform (LeCun et al., 1998)), learning rate with Adam (Kingma and Ba, 2015) \( \{1e{-4}, \ldots, 1e{-1}\} \), dropout rates \( \{0.2, 0.4, 0.5\} \), attention and recurrent units \( \{32, 64\} \), and batch-size \( \{1, 4, 8\} \) with early stopping criteria.

4 Results

<table>
<thead>
<tr>
<th>Book Content</th>
<th>Chunks</th>
<th>1000 sents</th>
<th>All</th>
</tr>
</thead>
<tbody>
<tr>
<td>Methods</td>
<td></td>
<td>ST</td>
<td>MT</td>
</tr>
<tr>
<td>Majority Class</td>
<td>0.506</td>
<td>0.506</td>
<td>0.506</td>
</tr>
<tr>
<td>SentiWordNet+SVM</td>
<td>0.582</td>
<td>0.610</td>
<td>-</td>
</tr>
<tr>
<td>NRC + SVM</td>
<td>0.526</td>
<td>0.597</td>
<td>0.541</td>
</tr>
<tr>
<td>NRC + SVM</td>
<td>0.537</td>
<td>0.590</td>
<td>0.577</td>
</tr>
<tr>
<td>NRC + SVM</td>
<td>0.587</td>
<td>0.576</td>
<td>0.595</td>
</tr>
<tr>
<td>NRC + SVM</td>
<td>0.611</td>
<td>0.586</td>
<td>0.597</td>
</tr>
<tr>
<td>Emotion Flow</td>
<td>0.632</td>
<td>0.643</td>
<td>0.650</td>
</tr>
<tr>
<td>Emotion Flow</td>
<td>0.612</td>
<td>0.639</td>
<td>0.640</td>
</tr>
<tr>
<td>Emotion Flow</td>
<td>0.630</td>
<td>0.657</td>
<td>0.662</td>
</tr>
<tr>
<td>Emotion Flow</td>
<td>0.656</td>
<td>0.666</td>
<td>0.674</td>
</tr>
</tbody>
</table>

Table 1: Weighted F1-scores for success classification in single task (ST) and multi task (MT) settings with varying chunk sequences when using all the book or only the first 1K sentences. \( *p < 0.05 \) (McNemar significance test between Emotion Flow (chunks 50, MT, All) and NRC+SVM (chunk 10, MT, All))

The results also show that the MT setting is beneficial over the ST setting, whether we consider the first 1K sentences or the entire books. This finding is akin to Maharjan et al. (2017). Similar to them, we suspect the auxiliary task of genre

---

\(^1\)https://www.gutenberg.org/

\(^2\)https://www.goodreads.com/
classification is acting as a regularizer.

Considering only the first 1K sentences of books may miss out important details, especially when the only input to the model is the distribution of emotions. Generally, the climax of the story is of utmost importance as it reveals the answers to the suspense, the events, and the emotional ups and downs in characters that build up through the course of the book. Accordingly, our results show that it is important to consider emotions from the entire book rather than from just the first 1K sentences.

5 Attention Analysis

From Figure 3, we see that using the attention mechanism to aggregate vectors is better than just concatenating the final forward and backward hidden states to represent the book in both ST and MT settings. We also observe that the MT setting mostly performs better than the ST setting.

Figure 4: Attention weights visualization per genre.

Figure 4 plots the heatmap of the average attention weights for test books grouped by their genre. The model has learned that the last two to three chunks that represent the climax, are most important for predicting success. Since this is a bidirectional RNN model, hidden representations for each chunk carry information from the whole book. Thus, using only the last chunks will probably result in lower performance. Also, the weights visualization shows an interesting pattern for Poetry. In Poetry, emotions are distributed across the different regions. This may be due to sudden important events or abrupt change in emotions. For Short stories, initial chunks also receive some weights, suggesting the importance of the premise.

6 Emotion Analysis

Climax Emotions: Since the last chunk is assigned more weights than other chunks, we used information gain to rank features of that chunk. From Figure 5, we see that features capturing the variation of different emotions are ranked higher than features capturing the average scores. This suggests that readers tend to enjoy emotional ups and downs portrayed in books, making the standard deviation features more important than the average features for the same emotions.

Table 2 shows the mean (\( \mu \)) and standard deviation (\( \sigma \)) for different emotions extracted for all the data, and further categorized by Successful and Unsuccessful label from the last chunk. We see that authors generally end books with higher rates of positive words (\( \mu = 0.888 \)) than negative words (\( \mu = 0.599 \)) and the difference is significant (\( p < 0.001 \)). Similarly the means for anticipation, joy, trust, and fear are higher than for sadness, surprise, anger, and disgust. This further validates that authors prefer happy ending. Moving on to Successful and Unsuccessful categories, we see that the means for Successful books
are higher than Unsuccessful books (highly significant, \( p < 0.001 \), for anger, anticipation, disgust, fear, joy, and sadness, and slightly significant, \( p < 0.02 \), for trust, and surprise). Moreover, the standard deviations for all emotions are significantly different across two categories (\( p < 0.001 \)). Thus, emotion concentration (\( \mu \)) and variation (\( \sigma \)) for Successful books are higher than for Unsuccessful books for all emotions in the NRC lexicon.

### Acknowledgments

We would like to thank the National Science Foundation for funding this work under award 1462141. We are also grateful to the anonymous reviewers for reviewing the paper and providing helpful comments and suggestions.

### References


Dzmitry Bahdanau, Kyunghyun Cho, and Yoshua Bengio. 2014. Neural machine translation by...


