Using Semantic Word Associations to Predict and Interpret the Success of Novels

Abstract

First published in 1868, Louisa May Alcott's Little Women, has never been out of print. Since its publication, approximately 1.78 million copies of Little Women have been sold, which equates to about 1,000 copies a month for 152 years. Every publisher in the industry hopes to find a manuscript that can sell even 100,000 copies, let alone 1,000 copies a month for 150 years. This begs the question: what makes Little Women a timeless success? Recently, and with some promising results, researchers have taken up the task of using machine learning and natural language processing to answer this question among others. In this paper, we attempt to predict a novel's success by modeling the lexical-semantic relationships of its contents. We built upon the previous research in this field and created the largest dataset used in such a project containing various lexical data from 18,000 books from Project Gutenberg. We utilized domain-specific feature reduction techniques to implement the most accurate models to date for predicting book success, with our best model achieving an average accuracy of 95.4%. While such strong performance in success prediction is impressive, we dug deeper to interpret the high accuracy. By analyzing the model parameters, we extracted the semantic relationships that separate successful books from the unsuccessful ones for books of 12 different genres. We found a mapping from WordNet's semantic word relations to a set of themes, as defined in *Roget's Thesaurus*. With this mapping, we discovered the themes that successful books of a given genre prioritize. In other words, if you want to write a bad children's book, write about keeping quiet in school.

CCS Concepts

• **Information systems** → **Data mining**; **Data extraction and integration**; *Content ranking*; • **Computing methodologies** → **Support vector machines**; *Cross-validation*; **Feature selection**; **Lexical semantics**; **Information extraction**; **Natural language processing**.

Keywords

book success prediction, semantic word association, feature reduction, book content mining

ACM Reference Format:

. 2021. Using Semantic Word Associations to Predict and Interpret the Success of Novels. In *The 14th ACM International Conference on Web Search and Data Mining, March 8–12, 2021, Jerusalem, Israel.* ACM, New York, NY, USA, 7 pages. https://doi.org/tbd

1 Introduction

Predicting the success of a novel by analyzing its content is a challenging research problem. Thousands of new books are published every year, and only a fraction of them achieve wide popularity. So the prediction of a book success could be exceptionally useful to the publishing industry and enable editors to make better decisions. Many factors contribute to a book's success including, but not limited to, plot, setting, character development, etc. Additionally, there are some other factors that contribute to a book's popularity that an author and publisher cannot control like the time when the book is published, the author's reputation, and the marketing strategy. In this paper, we only focus on the content of the book to predict its popularity.

Previous Work

The authors of [4] were the first to use statistical stylometry to predict the success of a novel based only on the contents of its first 1,000 sentences. Ashok et al. used stylistic approaches, such as unigram, bi-gram, distribution of the parts-of-speech, grammatical rules, constituent tags, and sentiment and connotation values as features with a Linear SVM [7] for the classification task. The authors used books from 8 total genres, and they were able to achieve an average accuracy of 75.7% for across all genres excluding HISTORICAL FICTION.

In [11], Maharajan et al. used a set of hand crafted features in combination with a recurrent neural network and generated feature

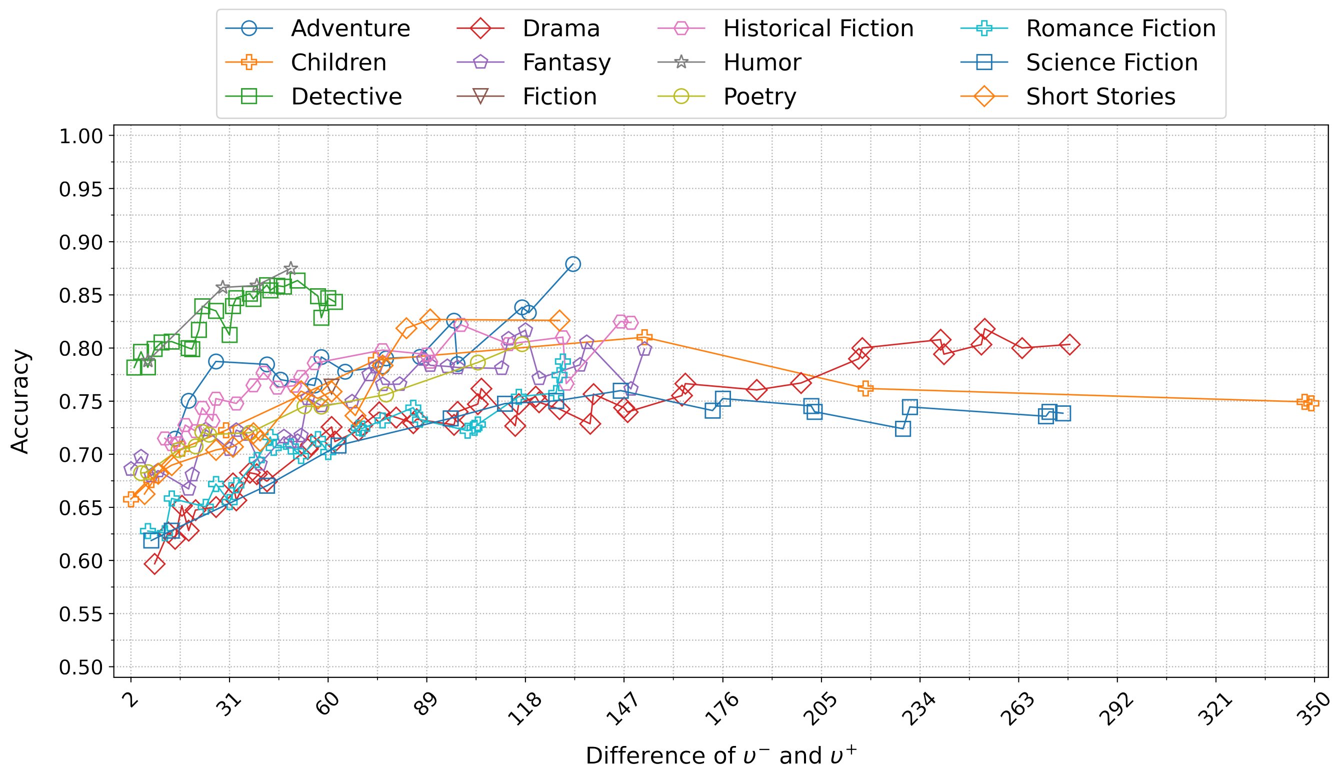
WSDM 2021, March 8–12, 2021 , Jerusalem, Israel

ACM ISBN tbd. . . $tbd
https://doi.org/tbd

Table 1: # of novels per genre and download count thresholds for unsuccessful ($\leq\ \upsilon^{-}$) and successful ($\geq\ \upsilon^{+}$) classes for the WordNet model

GENRE	# BOOKS	υ^{-}	υ^{+}
Adventure	917	11	143
Children	3278	7	160
Detective	285	34	85
Drama	785	12	265
Fantasy	382	45	163
Fiction	5369	11	72
Historical Fiction	961	10	156
Humor	1024	9	58
Poetry	1664	11	128
Romance Fiction	634	15	144
Science Fiction	1748	19	165
Short Stories	915	15	105
All	17,962	17	79

Figure 1: WordNet book success prediction accuracy by genre vs. difference of υ^- and υ^+



representation to predict the likelihood of novel success. The authors of [11] obtained an average accuracy of 73.5% for across 8 genres. They also performed several experiments, including using all the features used in [4], sentiment concepts [6], different readability metrics, doc2vec [10] representation of a book, and unaligned word2vec [14] model of the book.

In a more recent work [12], Maharajan et al. used the flow of the emotions across books for success prediction and obtained an F1-score of 69%. They divided the book into chunks, counted the frequency of emotional associations for each word using the NRC emotion lexicon [15], and then employed a recurrent neural network with an attention mechanism to predict both the genre and the success.

New Work and Improvements

We discovered various issues with the dataset used in [4][1] including, but not limited to its size, contents, and uniformity. This original dataset is quite small as it only includes the first 1,000 sentences from 800 books split into 8 different genres, which are further split into successful and unsuccessful classes, each having 50 books. Additionally, many of the files included have less than 1,000 sentences, or contain automatically generated text from Project Gutenberg instead of the text from the proper novel. Finally, the books included are prelabeled with their successful/unsuccessful class, which limits further testing.

[1] https://www3.cs.stonybrook.edu/~songfeng/success/

Considering these issues, we decided to build upon [4], but made following critical changes:

- Built the largest dataset containing a total of 17,962 books.
- Analyzed the *entire* content of each book and employed alternative prediction models.
- Introduced our feature reduction method to further improve model performance.

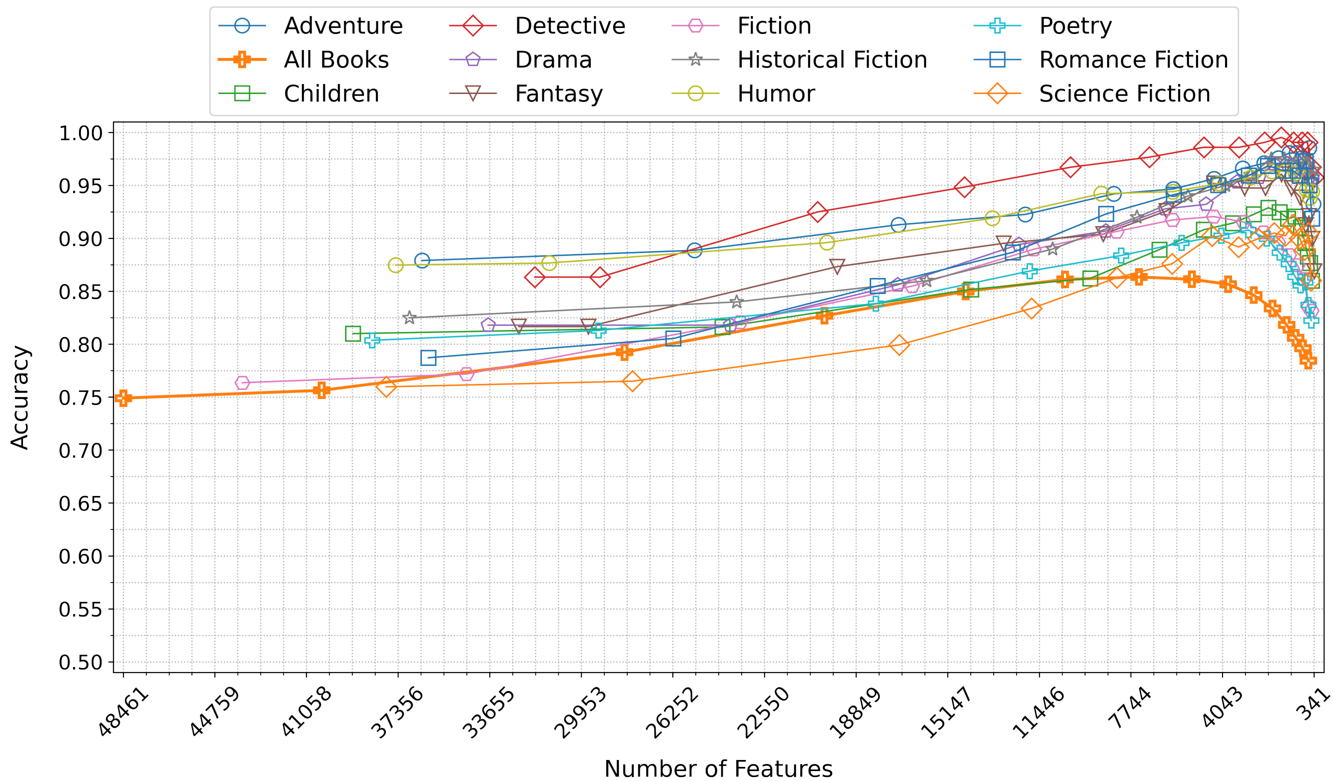
This leads to the motivation for this research and subsequent hypothesis. We believed that we could greatly improve upon the results of [4] with a cleaner and more complete dataset. We hypothesized that there must be at least one model that is both more accurate and more general than uni-gram, and from such a model, we could discover more interesting and revealing qualities that separate successful from non-successful books. Ultimately, through our improved methodology and larger dataset, our best models achieve over 95% accuracy for success prediction and identify the thematic elements prioritized by successful novels of a given genre.

2 Dataset Construction

We downloaded and used 17,962 English novels from Project Gutenberg. It is an online catalog of over 60,000 books, which are available to download for free in various formats [2]. We used a bash script to harvest the novels from Project Gutenberg according to the webmaster's guidelines [19][2].

[2] https://www.gutenberg.org/wiki/Gutenberg:Information_About_Robot_Access_to_our_Pages

Figure 2: Feature reduction process: WordNet success prediction accuracy vs. number of features



After downloading the books we used the NLTK API for data processing [5]. For each book, we extracted the uni-gram frequencies, the part-of-speech (POS) tag frequencies using the Stanford CoreNLPParser, the *Roget's Thesaurus* Category frequencies, and the WordNet Synset frequencies [18, 20, 21]. Like the authors of [12], we also extracted the NRC Emotional Lexicon features, and additionally the Linguistic Inquiry and Word Count (LIWC) features from each book. These emotional word mappings are highly valuable for some tasks, but the resulting models were not effective in our tests, and therefore not presented in this article [15, 17].

Each book in Project Gutenberg's catalog includes important information including its title, author, genre, language, and number of downloads. This metadata is available for download as RDF files[3]. We downloaded the RDF catalog, and then parsed and extracted the metadata for each book. For the experiments in this project, we use novels from 12 different genres as shown in Table 1.

Like in [4], we also used the download count of each book to define a measurement of success. In addition to predicting success classification for books split into unique genres, we also tested prediction performance independent of genre across the entire dataset. In both settings, we found an upper (υ^+) and lower (υ^-) download threshold for classifying books of that genre as "successful" or "not successful."

We performed an exhaustive search to find these thresholds by setting the class labels according to an incrementally widening download margin, and then training and testing each model at each increment. Starting at the median number of downloads, we label all books with downloads above the median as successful and all books below the median as unsuccessful. We train and test the given model with these labels and record the accuracy. Next, we move the upper bound to the first value greater than the median downloads and the lower bound to the first value less than the median downloads and train and test the model again with these new labels. This continues until there are less than 100 books in either class.

Each model achieved its best performance with a different class label margin, which we then used for the remainder of classification testing as shown for the WordNet model in Table 1, and the corresponding search plot in Figure 1. The search plot illustrates the lack of uniformity of download counts among and within each genre. At each pair of υ^+ and υ^-, the number of books in the successful and unsuccessful classes is not the same from genre to genre, just as the values for υ^+ and υ^- are not the same.

3 Methodology

Linguistic Models

We utilized six linguistic models for our quantitative analysis. Two of the models are our own implementation of models used in [4]. Our four additional models have not been used to make these types of

[3]https://www.gutenberg.org/wiki/Gutenberg:Feeds

Table 2: Accuracy (%) of classification results by genre, with/without feature reduction (R) (*best performance in bold*)

Model	Genre												Avg
	Adv.	Child.	Detective	Drama	Fantasy	Fiction	Hist. Fict.	Humor	Poetry	Romance	Sci-Fi	Short	
Unigram	77.0	77.2	84.2	74.7	78.4	70.2	76.8	84.2	74.0	73.4	76.0	69.6	76.3
Unigram^R	79.1	82.7	86.7	78.3	80.4	75.6	80.4	88.3	79.0	77.9	81.3	78.2	80.6
POS	79.7	76.4	77.2	69.1	75.4	73.7	84.9	86.7	77.6	72.9	74.9	79.3	77.3
POS^R	80.2	77.9	79.3	70.0	77.4	74.3	84.9	87.3	77.6	75.6	75.5	80.8	78.4
Roget	80.9	82.5	81.3	80.4	79.4	76.6	85.9	89.2	80.8	76.1	77.8	83.8	81.2
Roget^R	87.2	88.5	94.9	85.9	92.0	79.9	87.8	91.5	84.5	85.2	82.1	90.5	87.5
WordNet	87.9	81.0	86.3	81.8	81.7	76.4	82.5	87.5	80.4	78.7	76.0	82.7	81.9
WordNet^R	**98.5**	92.9	**99.5**	97.0	96.1	**92.1**	97.5	96.3	**90.8**	97.3	91.3	**95.4**	**95.4**
WNRC	94.3	89.8	92.4	92.8	86.0	82.1	97.0	94.4	82.4	91.9	84.9	86.4	89.5
WNRC^R	**98.5**	**93.7**	**99.5**	**98.3**	**96.9**	85.9	**100.0**	**97.3**	87.5	**97.7**	**93.9**	**95.4**	**95.4**
WNRT	80.7	77.7	82.5	82.3	73.8	74.0	93.0	87.7	77.6	85.5	78.6	84.5	81.5
WNRT^R	90.8	78.3	87.7	84.0	79.5	74.8	92.0	88.6	79.3	83.7	81.5	84.3	83.7

qualitative conclusions until now. These models include WordNet [5], *Roget's Thesaurus* [20], and two other models that map WordNet to different levels of *Roget's Thesaurus*.

I **Lexical Choices:** The words used in written documents is frequently employed for various applications, with the most popular lexical model being the n-gram model. For our analysis, we utilized the following lexical choice analysis models:
- **Unigram:** The frequency of unique words in the text.
- **WordNet:** WordNet is large lexical database of English words. The WordNet database groups nouns, verbs, adjectives, and adverbs into sets of cognitive synonyms called Synsets. Each Sysnet expresses a distinct concept and is represented by a single word. Since Sysnets represent conceptual synonyms, they are able to be linked through conceptual and semantic relationships [18]. WordNet has a total of 117,659 Synsets, each represented by a single, unique word, and our model uses the frequencies of these Synsets in each book. Not only does WordNet fit our semantic relation analysis methodology, but it has been used for the relevant task of metaphor identification in [13].
- ***Roget's Thesaurus*:** A tree structured thesaurus with six root nodes, which we will refer to as Roget Classes or Classes for short. Each Class is divided in sections, which results in 23 total sections. These sections represent 23 unique concepts that are both general enough to encompass a wide range of ideas, but also specific enough to retain clear meaning. Therefore, we refer to these sections as Themes and they are the critical piece to interpreting the results of class prediction. Themes are further divided into subsections, levels, etc. before terminating in 1,039 groups of synonyms, which we will refer to as Categories. The Categories are comprised of 56,769 total words, with about half appearing in multiple Categories [20]. Our Roget model uses the frequencies of these Categories in each book. Furthermore, the authors of [3] demonstrated the possible applications of *Roget's Thesaurus* for emotion detection with natural language processing, and [8] used the thesaurus for the related process of text summarizing.
- **Mapping WordNet to Roget:** Since *Roget's Thesaurus* has fewer synonym groups than WordNet (1,039 vs. 117,659), and those groups are hierarchically abstracted with each of the 1,039 Roget Categories belonging to one of the 23 Roget Themes, we mapped WordNet's Synsets to *Roget's Thesaurus* to discover more meaningful insights into the distinct characteristics of successful novels. We mapped WordNet to Roget Categories (WNRC), and then subsequently to Roget Themes (WNRT).

II **Part-of-Speech Distribution:** The authors of [4] demonstrated the value of POS tag distribution in success prediction, and [9] presented the relationship between POS tagging and genre detection and authorship attribution. Therefore, we reevaluated the application of POS tag distribution for success prediction.

Implementation

We used the sci-kit learn implementation of LibLinear SVM with 5-fold cross validation for class prediction [7, 16]. Part-of-speech tag features are scaled with unit normalization, while all other features are scaled using tf-idf. We used two strategies for all class prediction tasks:

- predicting class by genre, and
- predicting class independent of genre.

After the initial training and testing of each model, we employed an exhaustive feature reduction method, similar to our success labeling process, to maximize performance. For a given model, we start with the mean feature weight learned during training. We remove all features from the dataset with weights less than the mean feature weight. Next, we train and test the model on this reduced feature set and record the accuracy. For each subsequent test, starting at a step value of 0.25, we take only the features with weights greater than or equal to $Mean(OriginalWeights) + (StdDev(OriginalWeights) * Step)$. This process continues, increasing the step value by 0.25 after each iteration, until one of the following conditions is met:

- perfect accuracy is achieved,
- maximum accuracy is found (determined if consecutive subsequent feature sets produce decreasing performance), or

Table 3: Accuracy (%) of classification results for all books, before/after feature reduction

MODEL	ACCURACY	ACCURACY^R
Unigram	68.7	**70.3**
POS	67.4	67.4
Roget	73.5	74.8
WordNet	74.9	**86.3**
WNRC	76.4	**77.4**
WNRT	69.5	69.5

- the number of features is reduced to less than 1% of the original number of features.

4 Experimental Results

Classification Results

The prediction accuracy for each model by genre, and each model across all books, both before and after feature reduction are shown in Table 2 and Table 3, respectively[4]. As illustrated in both settings, the performance of nearly every model improved after we reduced the features with WordNet showing the largest improvement of an average of 11.9% when reduced by genre and 8.8% when reduced independent of genre.

The best performing models are indicated in bold in Table 2 and Table 3. When predicting novel success by genre, WordNet^R and WNRC^R show the best results with both models predicting a book's success class at 95.4%. WordNet^R is most accurate with ADVENTURE books, with an accuracy of 98.5%, while WNRC^R is able to predict the success of HISTORICAL FICTION books with 100% accuracy. When predicting the success of a book independent of genre, WordNet^R remains the most accurate at 86.3%.

Figure 2 illustrates the pattern of performance improvement that each model exhibits through the feature reduction process both by genre and independent of genre. As the number of features is reduced, the average accuracy for success prediction increases until the algorithm finds the best set of features and achieves peak performance. Then accuracy sharply drops as the feature set is reduced further. The fact that each model demonstrates such behavior validates the effectiveness of our feature reduction method.

Interpreting Book Success

While our reduced WordNet model displays excellent performance in both test settings (by genre and independent of genre), the resulting feature sets are not self-explanatory. In other words, the Synsets that the model deems most important do not necessarily highlight some interesting aspect of successful books, expected or otherwise. This is where *Roget's Thesaurus* proves most valuable.

We figured that if we looked up the Roget Theme of each WordNet Synset that we would find that the successful and unsuccessful books prioritize different Themes. This was possible due to the similarity in the structure of WordNet and *Roget's Thesaurus* as explained in the **Methodology** section above. With this hypothesis in mind, we mapped the reduced WordNet model to a new Roget model by first looking up the Roget Category of each Synset from the reduced WordNet feature set, and then summing the frequencies in each group of Sysnets. Then, as we did with each previous model, we reduced the new WordNet-to-Roget-Category (WNRC) model. From the WNRC^R model we mapped again, this time from Roget Categories to the 23 Roget Themes, which produced the WordNet-to-Roget-Themes (WNRT) model.

Table 4: Number of features before/after reduction for WordNet, WNRC, and WNRT

GENRE	WORDNET	WNRC	WNRT
Adventure	36,390	425	22
Adventure^R	540	149	17
Children	39,179	864	23
Children^R	2,117	210	14
Detective	31,833	840	21
Detective^R	1,670	272	9
Drama	33,718	812	23
Drama^R	1,996	333	3
Fantasy	32,483	779	21
Fantasy^R	1,665	192	12
Fiction	43,655	963	23
Fiction^R	4,407	403	10
Hist. Fict.	36,902	815	22
Hist. Fict.^{R}	864	198	14
Humor	37,457	822	22
Humor^R	1,493	324	7
Poetry	38,412	872	23
Poetry^R	3,004	243	7
Romance	36,137	451	23
Romance^R	688	129	12
Sci-Fi	37,835	617	23
Sci-Fi^R	1,190	225	6
Short	36,912	964	22
Short^R	3,769	217	16
All	48,461	1,001	23
All^R	7,415	374	23

We did not expect the performance of the WNRC model, since it was conceived strictly as an intermediary map between WordNet and Roget Themes. WNRC produced the highest baseline results of all the models used in our experiments with 89.5% average accuracy by genre, and was able to perfectly predict the success classification of HISTORICAL FICTION novels. Furthermore, WNRC^R accurately predicts success classification per genre at an average rate of 95.4%, tying it with WordNet^R as the best performing models we tested. What's impressive about the accuracy of WNRC^R when compared to that of WordNet^R is the large difference in number of features used in each model as shown in Table 4.

With such impressive results from WNRC^R, we expected WNRT and WNRT^R to follow suit despite learning with a feature set of at most 23 features. WNRT^R achieves an average accuracy of 83.7% learning from an average of only 10 features. While the performance of WNRT^R is impressive given the few number of features it requires, the purpose of WNRT^R was not to outperform WordNet^R or WNRC^R. As previously state, the motivation for the construction of

[4] All WordNet to Roget models are mapped from WordNet^R

Table 5: Top 5 most important themes for classifying CHILDREN novels and successful/unsuccessful thematic words

THEME	WORDS	
	Successful	Unsuccessful
Affections	enthusiastic, lively, tenderness	inactive, sluggish, dull
Communication of Ideas	secret, untruth, language	school, grammar, taciturnity
Formation of Ideas	incredulity, impossibility, curiosity	dissent, sanity, memory
Moral	gluttony, impurity, selfishness	punishment, virtue, duty
Personal	expecting, blemish, hopelessness	aggravation, dejection, dullness

WNRT was strictly to find a common thread between successful novels in each genre. That said, the decent performance of the WNRT^R model does support the reasoning behind its conception, and provide further evidence that WordNet^R and WNRC^R are general models that can reveal underlying characteristics of successful books.

Additionally, WNRT does not improve performance after feature reduction when classifying independent of genre. This outcome also supports our original hypothesis as it shows that the model requires each of the 23 Roget Themes in order to make the most accurate prediction. The lack of improvement in WNRT^R when predicting success class independent of genre also demonstrates the relationship between a novel's genre and its prioritization of certain Themes.

Successful Lexical Choices

After mapping the resulting feature weights of our WordNet^R model to Roget Themes, we were able to highlight the most important Themes when classifying the success of a novel given its genre. Table 5 gives the most important themes in predicting the success of CHILDREN'S novels and the successful and unsuccessful semantic word groups within those themes. These results clearly identify words associated with "school" and "grammar" as key contributors to unsuccessful CHILDREN'S novels, while words like "secret," "enthusiastic," and "selfishness" contribute to successful CHILDREN'S novels.

Table 6: Ranking the use of the most important CHILDREN'S themes for #1 downloaded CHILDREN'S book, *Little Women* relative to other CHILDREN'S books in the dataset

THEME	RANK
Communication of Ideas	2
Formation of Ideas	2
Personal	2
Moral	3
Affections	8

The indicated Themes align with intuitive expectations for CHILDREN'S books, especially the presence of FORMATION OF IDEAS and MORAL. To verify these results, we looked at the most downloaded CHILDREN'S book, *Little Women*. We ranked each book in the CHILDREN'S genre according to the frequency of each prioritized Themes listed in Table 6. Then, we looked to see where *Little Women* ranked for each of the Themes. *Little Women*'s use of the top Themes matches up as expected, as it ranks in the top three for four of the five most important Themes, and eighth for the fifth as shown in Table 6. The opposite is true for the least downloaded books, which all rank at the bottom for use of the most important Themes.

Our Thematic observations hold true for each genre, but there is not one theme shared by all 12 genres.This adheres to the observation we made about the WNRT and its lack of improvement after feature reduction for predicting success of all books.

5 Future Work

The discoveries made in our research are just the beginning of what can be done with our dataset. In addition to the data utilized for this project, we also extracted bi-gram and context-free grammar production features from each book. In future work, we will continue to explore the impact of these features in addition to semantic word associations on book success.

We also believe that we could achieve better results through the use of a different success surrogate metric. The scale of Project Gutenberg's catalog does not correspond to the website's popularity. Therefore, as we continue this work we will include ratings from popular sites such as Goodreads.com, Amazon, etc. to improve the our success labeling method [1].

6 Conclusion

We created the largest dataset for evaluating book success, and presented a novel study of semantic word association of a book's content to predict its success. Our empirical results demonstrate that semantic word association and its related thematic concepts are very useful in capturing a book's literary content and can predict a book's success with better accuracy. Rather than individual word frequency, the set of words with similar meaning has been proved to be most effective. The analysis performed in this project demonstrates the relationship between thematic word groups and a book's popularity, with our best models achieving a prediction accuracy of 95.4%. Finally, we illustrated that readers expect certain themes to be prioritized over others based on a book's genre, and the proper use of those themes directly contributes to a book's popularity.

References

[1] [n.d.]. *Goodreads*. https://www.goodreads.com/
[2] [n.d.]. *Project Gutenberg. (n.d.).* www.gutenberg.org.
[3] Saima Aman and Stan Szpakowicz. 2008. Using roget's thesaurus for fine-grained emotion recognition. In *Proceedings of the Third International Joint Conference on Natural Language Processing: Volume-I*.
[4] Vikas Ganjigunte Ashok, Song Feng, and Yejin Choi. 2013. Success with Style: Using Writing Style to Predict the Success of Novels. In *Proceedings of the 2013 Conference on Empirical Methods in Natural Language Processing*. Association for Computational Linguistics, Seattle, Washington, USA, 1753–1764. https://www.aclweb.org/anthology/D13-1181
[5] Steven Bird, Ewan Klein, and Edward Loper. 2009. *Natural Language Processing with Python*. O'Reilly Media Inc.

[6] Erik Cambria, Soujanya Poria, Devamanyu Hazarika, and Kenneth Kwok. 2018. SenticNet 5: Discovering conceptual primitives for sentiment analysis by means of context embeddings. In *Thirty-Second AAAI Conference on Artificial Intelligence*.

[7] Rong-En Fan, Kai-Wei Chang, Cho-Jui Hsieh, Xiang-Rui Wang, and Chih-Jen Lin. 2008. LIBLINEAR: A library for large linear classification. *Journal of machine learning research* 9, Aug (2008), 1871–1874.

[8] Alistair Kennedy and Stan Szpakowicz. 2010. Evaluation of a Sentence Ranker for Text Summarization Based on Roget's Thesaurus. In *Text, Speech and Dialogue*, Petr Sojka, Aleš Horák, Ivan Kopeček, and Karel Pala (Eds.). Springer Berlin Heidelberg, Berlin, Heidelberg, 101–108.

[9] Moshe Koppel, Jonathan Schler, Shlomo Argamon, and Eran Messeri. 2006. Authorship Attribution with Thousands of Candidate Authors. In *Proceedings of the 29th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval* (Seattle, Washington, USA) *(SIGIR '06)*. Association for Computing Machinery, New York, NY, USA, 659–660. https://doi.org/10.1145/1148170.1148304

[10] Quoc Le and Tomas Mikolov. 2014. Distributed representations of sentences and documents. In *International conference on machine learning*. 1188–1196.

[11] Suraj Maharjan, John Arevalo, Manuel Montes, Fabio A González, and Thamar Solorio. 2017. A multi-task approach to predict likability of books. In *Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics: Volume 1, Long Papers*. 1217–1227.

[12] Suraj Maharjan, Sudipta Kar, Manuel Montes-y Gómez, Fabio A Gonzalez, and Thamar Solorio. 2018. Letting emotions flow: Success prediction by modeling the flow of emotions in books. *arXiv preprint arXiv:1805.09746* (2018).

[13] Rui Mao, Chenghua Lin, and Frank Guerin. 2018. Word embedding and wordnet based metaphor identification and interpretation. In *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*. 1222–1231.

[14] Tomas Mikolov, Kai Chen, Gregory S Corrado, and Jeffrey A Dean. 2015. Computing numeric representations of words in a high-dimensional space. US Patent 9,037,464.

[15] Saif M. Mohammad and Peter D. Turney. 2013. Crowdsourcing a Word-Emotion Association Lexicon. 29, 3 (2013), 436–465.

[16] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay. 2011. Scikit-learn: Machine Learning in Python. *Journal of Machine Learning Research* 12 (2011), 2825–2830.

[17] James Pennebaker, Roger Booth, Ryan Boyd, and Martha Francis. 2015. Linguistic Inquiry and Word Count: LIWC2015. (09 2015).

[18] Princeton University. 2010. "About WordNet". https://wordnet.princeton.edu/ https://wordnet.princeton.edu/.

[19] Reason. [n.d.]. *How to Politely Download All English Language Text Format Files from Project Gutenberg*. https://www.exratione.com/2014/11/how-to-politely-download-all-english-language-text-format-files-from-project-gutenberg/

[20] Peter Mark Roget. 1852. *Roget's Thesaurus*. Project Gutenberg. http://www.gutenberg.org/files/10681/10681-h/10681-h.htm http://www.gutenberg.org/files/10681/10681-h/10681-h.htm.

[21] Muhua Zhu, Yue Zhang, Wenliang Chen, Min Zhang, and Jingbo Zhu. 2013. Fast and Accurate Shift-Reduce Constituent Parsing. https://doi.org/10.13140/2.1.2791.2961