Read to Grow: Exploring Metadata of Books to Make Intriguing Book Recommendations for Teenage Readers*

Yiu-Kai Ng

Computer Science Department, Brigham Young University, 3361 TMCB, Provo, Utah, 84602, USA.

Contributing authors: ng@cs.byu.edu;

Abstract

It is clearly established that spending time reading is beneficial for an individual's development in terms of their social, emotional, and intellectual capabilities. This is especially true for teenagers who are in the growing process and reading can improve their memory, vocabulary, concentration and attention span, creativity and imagination, and writing skills. With the overwhelming volume of (online) books available these days, it becomes a huge challenge to find suitable and appealing books to read. Current book recommender systems, however, do not adequately capitalize teenagers' specific needs such as readability levels, emotional capabilities, and subject's comprehension, that are more at the forefront for teenage readers than adults and children. To make appropriate recommendations on books for teenagers, we propose a book recommender system, called TBRec. TBRec recommends books to teenagers based on their personal preferences and needs that are determined by using various book features. These features, which include book genres, topic relevance, emotion traits, readers' advisory, predicted user rating, and readability level, have significant impact on the teenagers' preference and satisfaction on a book. These distinguished parts of a book, which are premeditated and essential criteria for book selection, identify the type, subject area, state of consciousness, appeal factors, (un)likeness, and complexity of the book content, respectively. Experimental results reveal that TBRec outperforms Amazon, Barnes & Noble, and LibraryThing, three of the widely-used book recommenders, in making book recommendations for teenagers, and the results are statistically significant.

Keywords: Teenagers, books, recommender systems, metadata

1 Introduction

Reading proficiency has a large impact on both personal growth and how an individual contributes and participates in the wider society. On the individual level, successful readers benefit from improved self-esteem and positive self-concept [69]. Furthermore, reading books is associated with "better health, volunteering, and strong satisfaction with life" [29]. Reading has also been found to help develop social skills and overall social competence [69]. For an individual to participate in the larger society, they need to cooperate well. Cooperation is only made possible by effective communication, which can directly be improved by reading ability and has been linked with increased writing ability [16]. The teenage years are an excellent stage for individuals to put a focused effort into improving these personal skills. As social media technologies become more widespread, many teenagers struggle with self-esteem and their concept of self. Improving reading skills would serve as an effective counter. Second, reading and writing becomes directly applicable in teenage years when school courses become more challenging and as individuals begin employment in later years. Adolescents can use long-term planning for goals of self-improvement and have fewer constraints on their time than in adulthood.

For the highest likelihood to reap all the benefits of reading, it is valuable to select appealing books to read. First and foremost, the more interesting a book is, the more likely the individual will want to engage with it. In adolescence, teenagers exert a concerted effort to find themselves, learn independence from their parents, and want to read books that are interesting to them. Moreover, it has been found that with increased interest, there comes increased growth. Guthrie et al. [27] show that reading involvement and interest assessed by in-depth interviews "significantly contributed to the prediction of reading comprehension growth". Hence, reading interest is important not only for frequency of engagement, but the quality of engagement and the learning outcomes as well. However, it is often difficult to select a book of interest. Individuals may have many different preferences and opinions in books that they would like to read. Books can indeed vary in topic, length, language, reading level, and style, just to name a few factors. The web search does allow for a larger selection, which makes it possible for readers to refine their selections and raise their standards of interest. Web search, however, is a tedious and time consuming process, due to the huge volume of online books. In 2019, there were 4.5 million ebooks on the Kindle store, while in May of 2021, there were an estimated 11 million Kindle ebooks [38].

To provide a good user searching and reading experience, we have developed a new book recommender system, called *TBRec*, that is designed specifically for and tailored to teenagers. TBRec greatly reduces the number of books for the user to browse through, primarily through filtering out choices deemed to be non-relevant. It is crucial for the user to have a choice, but studies have shown that limiting the choices increases the likelihood of selection, and the subsequent average satisfaction with the choice [32]. TBRec is unique, since it considers and incorporates a wide variety of features of books to suggest

appealing books to teenagers to enhance their reading experience. A number of these features, which include the genres [67], topic relevance [55], emotional trait [48], appeal terms [14], predicted ratings [1], and readability level of books [19], target this particular group of readers [13, 42] and are essential in making book recommendations for young readers [51, 52]. By integrating these features of a book, TBRec can offer recommendations that meet the "needs" and "tastes" of the targeted teenage users. Existing book recommenders for adults [3] and children [45], however, tailor especially for older and younger groups of readers, respectively with different requirements and expectations that are incompatible with teenage readers. TBRec is designed exclusively for teenagers that helps them develop a healthy reading habit. Clinical studies [18, 64] have indicated that there are significant differences between children and teenage readers in terms of their ability towards reading, reading scores, and verbal IQ that occur as early as first grade and the trajectory of these differences increases from childhood to adolescence.

2 Related Work

There exist a number of book recommender systems that adapt various machine learning or information retrieval approaches to make suggestions on books. Rather than relying on the history-based approach that suffers from the cold-start problem, Shu et al. [61] propose a content-based recommendation algorithm through a convoluted neural network. Constructing the required content profiles, however, is difficult, since they depend upon different types of user content, such as text, audio, video, and images, that are not always available. Putri and Zulkarnain [54] attempt to overcome the cold-start problem in making recommendations on academic-related materials by utilizing a word-embedding model as a topic model for their content-based recommendation system.

Zuo et al. [74] introduce a collaborative-filtering-based book recommender that considers social-networking system attributes to determine users' neighbors. Nearest neighbor approaches, however, are difficult to adopt when the required dataset is sparse and often run into the cold-start problem in determining user ratings on new books. Alharthi et al. [2] also use social media to determine interests of the user to make book recommendations. They specifically look into using tweets posted by users to "warm up" their account of recommendations.

Xin et al. [72] make recommendations for teenagers on published articles, instead of books, to read based on the related (sibling) categories of the articles. Liang et al. [40], on the other hand, leverage the reading patterns of users to estimate users' preference levels in making recommendations on online articles. The authors analyze a user's post-click behavior, i.e., mouse behavior, keyboard event, and page scrolling event, to determine his preference in online articles.

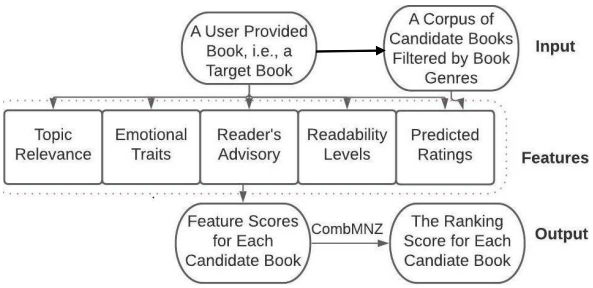


Fig. 1 The overall process of TBRec, the proposed book recommender for teenage readers

To reduce the data processing load in making book recommendations, Yingwei Zhou [73] develops a data mining approach with an apriori algorithm. The author mines the strong association rules and determines confidence coefficient to the processed data and by means of the improved Apriori data mining algorithm to generate association rule database for book suggestions.

TBRec is different from existing book recommenders, since the latter either recommend books for targeted audience other than teenagers that cannot be seamlessly adaptable for teenagers, or exclude the usage of book features that are essential in making appealing/desirable recommendations for teenagers.

3 A Teenager Book Recommender

TBRec considers essential book features to address the reading preferences and needs of teenagers in making appealing book recommendations to them. These features include *genres, topic relevance, emotion traits, reader's advisory, predicted user rating*, and *readability Level*, which analyze the category, subject area, sensibility feeling, reading preference, (dis)likeness, and comprehensive level of a book, respectively that are applicable to teenage readers [20, 51, 52, 63].

To assess and ascertain that recommended books are preferred and of interest to a user, TBRec requests the user to provide a book, called *target book*, that the user enjoyed reading. Using the target book Tk and a corpus C of published books, TBRec chooses a subset of books in C, called *candidate books*, based on the *degree of similarity* on genres assigned to Tk and books in C. Hereafter, TBRec proceeds to analyze the five different book features and obtain five different feature scores for each candidate book CB with respect to Tk. The closer the specific features of CB compared with Tk are, the higher the combined feature score of CB is. To incorporate the five feature scores into a final ranking score for CB, TBRec uses a simple, widely-used linear combination model, called *CombMNZ* [39], a well-established data fusion method without using optimal weights, i.e., all the involved features are weighted equally. Based on the final ranking scores of the candidate books, TBRec recommends them to the user in a ranked order. Figure 1 depicts the overall process of TBRec.

3.1 Filtering Candidate Books Based on Genres

A genre [23] of a book indicates a particular *category* of the book. The basic assumption of genres is that if two books have the *same* genre, then they are likely *similar* in terms of their plots and contents [26]. *Genres* for books come with various properties: (i) two different genres can be highly similar, such as "dark" and "mystery", (ii) a book is assigned different genres by different experts/users, and (iii) some books have multiple genres assigned to them. *Similarities* of book genres can be determined using *word-correlation factors* (defined in Section 3.1.1), and TBRec computes the score of the genres, denoted GS, of a book in a corpus (as detailed in Section 3.1.2) to determine its candidacy of a target book.

3.1.1 Word-Correlation Factors

To determine the similarity of two genres, we use the word-correlation factors in our *word-similarity* matrix, denoted $WS\text{-}matrix$, which is a 54,625 × 54,625 symmetric matrix. The similarity, denoted $Word\_Sim$, of any two non-stop, stemmed words i and j in $WS\text{-}matrix$ is computed by using the (i) *frequency* of co-occurrence and (ii) relative *distances* of i and j in each document in which they co-occur (as shown in Equation 1). $WS\text{-}matrix$ was constructed using the documents in the Wikipedia collection [70] with 930,000 documents written by more than 89,000 authors on various topics and writing styles.

$$Word\_Sim(i,j) = \frac{\sum_{D \in Wiki} \left(\frac{\sum_{k_i \in D} \sum_{k_j \in D} \frac{1}{d(k_i,k_j)+1}}{N_i \times N_j} \right)}{\mid Wiki \mid} \tag{1}$$

where $\mid Wiki \mid$ is the number of documents in the Wikipedia collection, i.e., $Wiki$, $d(k_i, k_j)$ denotes the *distance* (i.e., the number of words in) between words i and j or their stems in a Wiki document D in which they co-occur, and N_i (N_j, respectively) is the number of times word i (j, respectively) and its *stems* appeared in D.

Compared with synonyms and related words compiled by WordNet [71] in which each pair of words is not assigned a *similarity weight*, $Word\_Sim$ offers a more sophisticated measure of word similarity. Moreover, it has been shown that $Word\_Sim$ outperforms pre-trained word embeddings, a widely-used approach for solving NLP problems, and user tags in terms of measuring the degrees of similarity among different words for web search [50].

3.1.2 Computing Book Genre Scores

A book is assigned a number of genres to identify the category of the book. For example, the genres of the book "The Hunger Games" by Suzanne Collins include *Young Adult, Fiction, Fantasy, Romance, Adventure, Action*, and *Apocalyptic*. Each of these genres is also associated with the number of users who have chosen the label to identify the category of the book, as

Book	Horror	Thriller	Action	Adventure
1	50	40	6	4
2	20	9	8	4

Word_Sim(Horror, Thriller) = 0.93; *Word_Sim*(Horror, Action) = 0.35
Word_Sim(Horror, Adventure) = 0.15; *Word_Sim*(Thriller, Action) = 0.67
Word_Sim(Thriller, Adventure) = 0.33; *Word_Sim*(Action, Adventure) = 0.91
GSG(Horror, 2, 1) = 20/41 x 50/100 + (20/41 x 40/100 x *WS*(Horror, Thriller) + 20/41 x 6/100 x *WS*(Horror, Action) + 20/41 x 4/100 x *WS*(Horror, Adventure)) = **0.47**
GSG(Action, 2, 1) = 8/41 x 6/100 + (8/41 x 50/100 x *WS*(Action, Horror) + 8/41 x 40/100 x *WS*(Action, Thriller) + 8/41 x 4/100 x *WS*(Action, Adventure)) = **0.11**

Fig. 2 *GSG* scores computed for a corpus book, Book 2, based on genres in Book 2 and Book 1, a target book, and **Word_Sim** (WS)

in the Goodreads dataset [24]. For a corpus book to be qualified as a candidate book, denoted CB, of a *target book*, denoted Tk, it must (i) share at least a *common* genre with Tk, and (ii) its *genre score*, with respect to Tk, denoted $GS(CB, Tk)$, must be greater than the *genre score* of Tk itself, denoted $GS(Tk, Tk)$. TBRec considers the *top-10* ranked genres (based on their numbers of users who assigned the genres) in computing the two scores.

$$
\begin{aligned}
GS(CB, Tk) &= \frac{\sum_{G \in GSet_{CB}} GSG(G, CB, Tk)}{\mid GSet_{CB} \mid}, \text{where} \\
GSG(G, CB, Tk) &= \frac{\#\_Users(G)\ in\ CB}{\#\_Users(CB)} \times \frac{\#\_Users(G)\ in\ Tk}{\#\_Users(Tk)} + \\
&\sum_{j=1, G_j \neq G}^{|GSet_{CB}|} \frac{\#\_Users(G)\ in\ CB}{\#\_Users(CB)} \times GST(G_j, Tk) \times\ WS(G, G_j), \text{where} \\
GST(G_j, Tk) &= \begin{cases} \frac{\#\_Users(G_j)\ in\ Tk}{\#\_Users(Tk)} & \text{if } \#\_Users(G_j)\ in\ Tk > 0 \\ \frac{1}{\#\_Users(Tk)} & \text{Otherwise} \end{cases}
\end{aligned} \tag{2}
$$

where $GSet_{CB}$ is the set of genres in CB, which is 10 in our case, $WS(G, G_j)$, i.e., $Word\_Sim(G, G_j)$, is the *genre similarity* of G and G_j, $GS(Tk, Tk)$ is defined using the equation $GS(CB, Tk)$ by substituting CB by Tk, and $Word\_\ Sim(G, G_j) = Word\_Sim(G_j, G)$.

Equation 2 assigns the highest GSG to a genre G of CB if G in CB is the most dominated genre in CB. The second multiplication in GSG is used to assign weight to genres such that the *less similar* a genre G_j is to G, the *less* it is *weighted* in the computation of $GSG(G, CB, Tk)$. Figure 2 shows an example of applying Equation 2 to two different genres in two books, i.e., Books 1 and 2, which illustrates how the genre 'Horror' receives a higher score than the genre 'Action', since the genre 'Horror', along with the genre 'Thriller' that is highly similar to 'Horror', have higher number of labeled users than other genres.

We realize that solely relying on the genres of a candidate book CB to make recommendations could yield less-than-ideal results, since (i) a *large* amount of data are required to accurately determine a particular user's genre preferences, (ii) a lower (high, respectively) *genre similarity score* of CB does not necessarily indicate that a user dislikes (likes, respectively) CB, and (iii) books can be highly similar with respect to their genres and being dissimilar with respect to other features not related to genres. Therefore, TBRec considers

other features, i.e., topic relevance, emotion traits, readers' advisory, predicted user rating, and readability level, of CB in making book recommendations.

3.2 Topic Relevance

TBRec analyzes the topic, i.e., subject area, of a candidate book by using topic modeling to determine the most dominated topic covered in the book. Latent Dirichlet Allocation (LDA) model is one of the most popular topic modeling methods [5]. LDA generates a collection of topics by using a set of training text documents. Each topic is a list of keywords arranged by frequency of occurrence. Using the generated list of topics, a text document can be labeled by the LDA model with the possibility for each topic. TBRec adapts the LDA model to determine the topic of a candidate book.

3.2.1 Training a LDA Model

To train a LDA model there are two steps involved: (i) a series of documents to be analyzed for different topics, and (ii) an ideal number of latent topics to be generated. TBRec uses the *descriptions* of candidate books as the set of documents for creating the topics. We extracted 18,000 book descriptions from Goodreads, which is publicly available, as our training instances. Each one of the book descriptions consists of a sequence of words. During the training process, we tried different topic numbers between 5 and 40 and chose 20 topics as the ideal topic numbers, since 20 topics yield the highest relevance scores among the keywords within each topic than other topic number.

During the training process, LDA estimates the probability of a non-stop, stemmed word w given a (latent) topic z, i.e., $P(w \mid z)$, and the probability of a topic z given a text document D, i.e., a book description in our case, i.e., $P(z \mid D)$. A number of algorithms have been proposed for estimating $P(w \mid z)$ and $P(z \mid D)$, such as variational Bayes [36], expectation propagation [46], and Gibbs sampling [21]. We have chosen Gibbs sampling, since it is easier to implement, more efficient, faster to obtain good approximations, and easily extended than others [53].

3.2.2 Book Classification Using LDA

The classification process of LDA on a given candidate book CB can be described as finding the probabilities of a number of topics covered in the description D of CB and selecting the *topic* with the *highest probability* as the topic covered in D. After creating the latent topics on D, Equation 3 is employed to determine the selected topic of D based on the distribution of each non-stop, stemmed word in D and their probabilities in each topic z_j, i.e., $P(w_i \mid z_j)$, such that the topic z_j that has the *highest probability*, i.e., $P(z_j \mid D)$, is selected as the topic of D.

$$Topic\_Match(CB, D) = P(z_j \mid D) = \max_{j=1}^{N} \sum_{i=1}^{M} P(w_i \mid z_j) \tag{3}$$

where w_i is the i^{th} distinct word in D, M is the total number of distinct words in D, z_j is the j^{th} latent topic among all the N (= 20) latent topics, and is an input to LDA, and $P(w_i \mid z_j)$ is the *probability* of w_i in z_j.

Table 1 shows three of the 20 topics created for the books using book descriptions extracted from Goodreads.com. The books, "The Light Fantastic" and "Blood of the Fold", belong to the same topic, which is determined by using our trained LDA model, and some of the overlapped topic keywords exist in the descriptions of the two books are **highlighted** in Table 2.

Table 1 Three (out of the 20) topics created by using the trained LDA model for our teenage book recommender

Topic	Keywords
1	story, tale, reader, author, write, classic, great, collection, ...
2	make, game, good, run, job, money, show, business, company, fast, ...
20	man, turn, stop, catch, dangerous, deadline, plan, save, steal, black, ...

Table 2 Portions of the two sample book descriptions that are assigned to the same topic by the trained LDA model

Title	Sample Common Keywords
Light Fantastic	The hero happens to be the **wizard** Rincewind who was seen falling off the edge of the **world**
Blood of the Fold	Richard comes to terms with his true identity as a war **wizard** of the new **world**

3.3 Emotion Traits

Besides considering the topics of books for suggesting appealing books to be recommended to teenagers, TBRec also analyzes the emotional contents of a book [47, 57] and using them to compare the similarity of books and predict the likelihood a teenager will enjoy it based on their age. By utilizing trends apparent in age groups regarding emotional preferences, TBRec can recommend books with relevant, age-appropriate emotional compositions and focuses.

The study of basic domain of interpersonal relations and personality involves with eight basic emotions of a person, i.e., *anger, anticipation, disgust, fear, joy, sadness, surprise*, and *trust*, which are defined as the "primary emotion dimensions" [52]. Every *word* is associated with a combination of these eight fundamental emotions, and a *word* in a book content has a defined value in each emotion, which by default is zero. For example, for the word "death", all of its emotion values are defined except *Trust* and *Joy*. Death's highest-scoring emotion value, which is *Sadness*, is 0.915, while its lowest-scoring emotion value, *Anticipation*, is 0.398. The emotion values of each word in a book content can be used for capturing the emotion trait of the corresponding book.

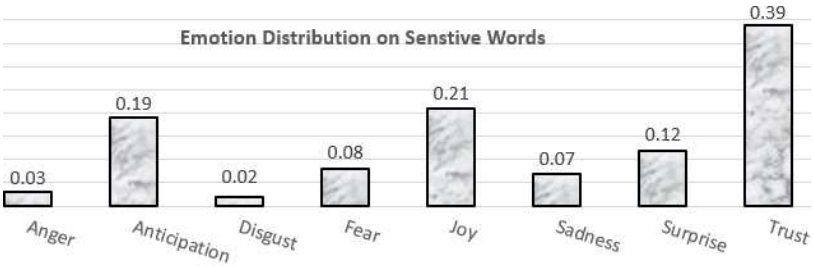


Fig. 3 The emotion values of sentiment words in the content of the book "Walk Two Moons" by Sharon Creech

Different books express different emotions and invoke readers with different feelings while reading. (Figure 3 depicts the emotion values of sentiment words in a book weaving on two tales, one funny and one bittersweet for youngsters.) TBRec measures book similarity partially by analyzing various emotions expressed in a book using a vector representation of the sentiment words expressed in the book content. A vector of normalized emotion association scores of the eight emotions is then constructed. The closer the emotion traits expressed between a candidate Book CB and the target book provided by a user U is, the more likely CB is appealing to U. The key factor of our emotion analysis is to assign each emotion value to a word in a book content to capture the emotions expressed by it.

As book content are composed of many (non-)sentiment words [17], we adapt the Emotion Intensity Lexicon (NRC-EIL) [9], developed by The National Research Council Canada (NRC), to evaluate their emotional values. NRC-EIL consists of 10,000 English words with real-valued levels of intensities for the eight emotions. NRC-EIL assigns eight emotion scores between 0 and 1, inclusively, to each non-stop, lemmatized word, and each of these scores becomes the component value of a unit vector that represents the emotion captured by the entire content.

TBRec calculates the frequency of each word with emotions, which is a non-stop, stemmed words in the book content of CB to obtain the combined association score for each emotion. As it turns out, 40-60% of words in a book content, called *Objective*, with emotion but are not accounted for in NRC-EIL and thus were initially excluded in an emotion vector. To improve the quality of the emotion vectors, we reduce the amount of words that are marked as Objective, the same technique adapted by Milton et al. [44]. TBRec utilizes the python package *synset* in the NRC-EIL which allows us to get a word's synonyms that we can use to search the NRC-EIL for their emotions in addition to the word itself. TBRec ends up checking all of the synonyms provided by synset for each Objective which yields more accurate emotion trait analysis results that are statistically significant ($p < 0.03$ on the t-test) compared with excluding the usage of synset on Objective. Figure 4 compares the emotion distribution on the sentiment words, including synonyms of Objective words, in the two books, a target book "Hyperion" and a candidate book "Rosemary's

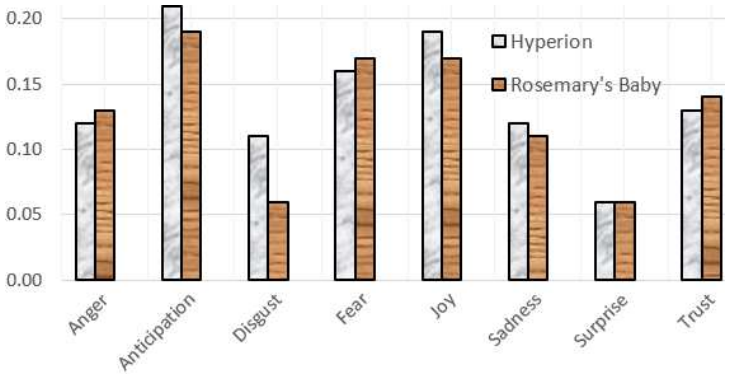


Fig. 4 Comparison of the emotion traits expressed in the sentiment words in two books, "Hyperion" by Dan Simmons and "Rosemary's Baby" by Ira Levin

Baby" and their cosine similarity measure yields the value of 0.98, which is highly similar.

3.4 Reader's Advisory

To motivate teenagers to read, it is necessary to avoid presenting these readers with books that are either too easy or too difficult to read or involve topics unappealing to them which could diminish their interest in reading [4]. In fact, finding the right books for the right audience is not easy. Even though existing book recommenders can assist ordinary readers in finding books of interest, they rely heavily on large historical data, e.g., personal tags, which might not be available on social media sites due to teenage privacy issues of readers. For this reason, TBRec applies reader's advisory in assisting teenagers to find appealing and age-appropriate books to read.

Reader's advisory (RA) helps users find books to read based on their reading preferences, a job common to most librarians. TBRec emulates the readers' advisory service, which has been available at public libraries since the late 1800's [59]. Reader's advisory offers (non-)fiction materials of potential interest with "the help of knowledgeable and non-judgmental library staff" [59]. While the traditional Reader's advisory model involves face-to-face discussions between patrons and librarians, these days existing technologies replace human interactions with online forms filled out by patrons to capture their interests. Besides analyzing the *topical areas* and *content descriptions* of books favored by a reader, during the RA process, librarians examine the *appeal factors* of books that the reader is interested in.

Librarians consider literary elements during the search process, where appeal factors are considered as part of literary elements [51]. Based on the appeal terms[1] that describe the appeal factors of books *preferred* by a reader,

[1] Appeal terms are different from *tags* created by common users of social media websites, since the latter can be inaccurate, noisy, or ambiguous.

Table 3 Appeal factors and corresponding appeal terms in Reader's Advisory

Appeal Factor	Appeal Terms
Frame	Bittersweet, contemporary, descriptive, upbeat, school, ...
Tone	Dark, happy, quirky, surreal, ...
Storyline	Action-oriented, character-centered, humorous, gentle, ...
Characterization	Believable, distant, dramatic, well-developed, ...
Language and Writing Style	Candid, complex, conversational, extravagant, poetic, prosaic, simple, ...
Pacing	Easy, fast, slow, ...
Special Topics	Addiction, bullying, illustration, violence, ...

librarians suggest other books matching (to a certain degree) the interests/preferences of the reader. However, due to the amount of books being published on a regular basis these days, it is an impossible task for a librarian to be familiar with every existing book to determine if it could be a potential relevant recommendation for the reader. For this reason, librarians turn to RA databases, which are available at NoveList, Fiction Connection, and Readers' Advisory Online, to conduct fact-based, appeal factor-oriented, and read-alike searches in locating books to suggest to a reader. (Table 3 shows the appeal factors and a sample of their corresponding appeal terms.) TBRec automates the reader's advisory to measure the similarity between the target book and a candidate book. We adapt the seven most significant appeal factors, i.e., *frame, tone, storyline, special topics, characterization, language* and *writing style*, and *pacing*, as appeal factors identified in [51]. Each appeal factor is associated with the corresponding appeal terms to represent the factor.

TBRec creates the appeal-term description for a book based on not only the appeal terms extracted from *user reviews* on the book, but also their *frequency of occurrence*. The latter captures the relative *degree of significance* of an appeal term in describing its corresponding factor based on reviewers' varied opinions on appeal factors that apply to the book. (A sample of the appeal term description generated using RA for a book is shown in Figure 5.)

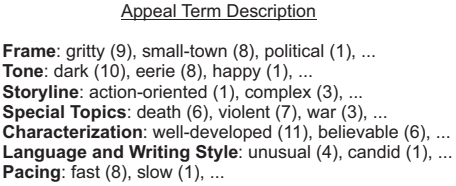


Fig. 5 RA-generated appeal-term description for "The Hunger Games", where the number indicates the *frequency* a term in describing its appeal factor in the reviews

TBRec determines the appeal factors and appeal terms by analyzing the distribution of words in user reviews, regardless of their length, for each candidate book. We consider a list of non-stop, stemmed words extracted from the user reviews of a candidate book and use the list to calculate the frequency of occurrence of each appeal term in each appeal factor. Hereafter, the literary

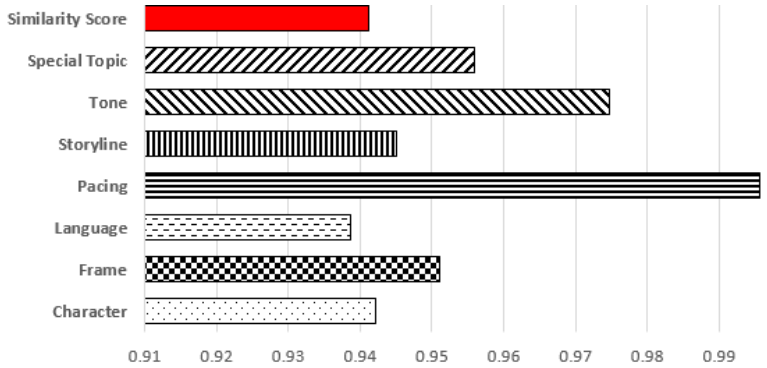


Fig. 6 The similarity score of the appeal terms-appeal factors in the reader's advisory on two teenage books, "Lonesome Dove" and "Peace Like a River"

elements of each book will be represented by seven vectors, each one representing the frequency of occurrence of each appeal term in each one of the seven appeal factors. Afterward, TBRec applies the cosine similarity measure on the vectors to calculate the degree of similarity of reader's advisory between the target book Tk and a candidate book CB, denoted $RASim(Tk, CB)$, using Equation 4.

$$RASim(Tk, CB) = \sum_{f \in F} \frac{Cos(\overrightarrow{Tk_f}, \overrightarrow{CB_f})}{\mid F \mid} = \sum_{f \in F} \frac{\frac{\overrightarrow{B_f} \cdot \overrightarrow{C_f}}{|\overrightarrow{B_f}| \times |\overrightarrow{C_f}|}}{\mid F \mid} \tag{4}$$

where F is the set of appeal factors and f is an appeal factor in F. The higher the reader's $RASim$ score is, the more likely the user would enjoy reading the candidate book CB given Tk.

Figure 6 depicts the overall and individual similarity score of the reader's advisory appeal factors on two teenager books, "Lonesome Dove" and "Peace Like a River". The former is an American classic novel of the American West that follows two aging Texas Rangers embarking on one last adventure, whereas the latter tells the story of an 11-year-old, asthmatic boy who has reason to believe in miracles. The similarity score of the two books shows that they are closely related in each of the appeal factor-appeal term comparisons, since all of the them are in the 90%.

3.5 User Rating Prediction

Rating prediction is a classical approach for making recommendations. Attempts have been made in the past by relating users to similar users and an item to similar items on which user- and item-based rating prediction systems have been developed. TBRec adapts the item-based collaborative filtering (CF) approach to predict the rating of a candidate book for a target book so that the higher a predicted rating on an item I for a user u using the ratings of items previously encountered by u is, the more likely I appeals to u.

This recommendation strategy is intuitive and relatively simple to implement. Moreover, it requires no costly training phases that are needed for machine learning models and thus is scalable to millions of users and items. In addition, it is not significantly affected by the constant addition of users, items, and ratings in a large number of commercial applications [37] and does not require retraining for each addition [6].

Based on the item-based CF approach, TBRec predicts the rating of a candidate book i for a target book by analyzing books previously rated by the target user u that are similar to i, denoted $N_u(i)$. Moreover, TBRec extends the item-based CF approach by determining the *degree of similarity* between i and each book j in $N_u(i)$, denoted $w_{i,j}$, to further enhance the item-based CF approach. TBRec applies the pre-defined word vectors, which capture the essential (i.e., non-stop, stemmed) keywords in the book descriptions of i and each book j in $N_u(i)$, to compute the (cosine) *similarity* between i and each book j in $N_u(i)$. The predicted user rating of i for user u, denoted $r_{u,i}$, is computed as follows.

$$r_{u,i} = \frac{\sum_{j \in N_u(i)} w_{i,j} \times r_{u,j}}{\sum_{j \in N_u(i)} w_{i,j}} \tag{5}$$

where $r_{u,j}$ is the rating provided by u on book j:

3.6 Readability Level Analysis

The readability level of a book is a useful measure for teenagers to identify reading materials suitable at their reading levels. The majority of published books, however, are assigned a readability level range, such as 8-12, by professionals, instead of a single readability level for their intended readers, which is not useful to the end-users who look for books at a particular grade level. This leads to the development of readability formulas/analysis tools [31, 35, 62, 63]. Unfortunately, these tools simply perform a one-dimensional analysis on a text based on shallow features, such as the average number of syllables per word (words per sentence, respectively), the average sentence length, and vocabulary lists, which might not precisely capture the complexity of a text [15]. Thus, to alleviate this constraint imposed on readability analysis on books, we have developed our own readability level analysis tool, called *Read_Level*, which relies heavily on metadata of books that is publicly and readily accessible from reputable book-affiliated online sources, besides using previews of books[2], to predict the readability level of books as detailed below.

3.6.1 Prediction Based on Book Textual Features

Read_Level considers *vocabulary* and the *count of syllables* as displayed in a book preview for partially predicting its readability level, which are also

[2]Previews of books can be extracted from the Book Cave dataset, which consists of more than 20,000 teenager books that are made available by publishers to showcase their books.

commonly used by traditional readability formulas. However, Read_ Level does not rely solely on any particular existing readability formula. To predict the readability level of a book, Read_Level considers seven *textual features* based on the count of (i) long words (with more than six letters), (ii) sentences, (iii) total words, (iv) letters, (v) syllables, (vi) words with three or more syllables, and (vii) unique unfamiliar words [63]. Since the length of the text, i.e., the total number of characters, available in a preview is different for each book, we normalize these counts to the length of the preview.

3.6.2 Grammar Prediction

Read_Level examines grammatical constructions, as defined by the US curriculum, to compute the values of grammar predictors. These predictors reflect the *complexity* of the (i) writing style, (ii) organization of the sentences, and (iii) grammatical constructs found in a text. The analysis of the grammar of textual content in a book (in the Book Cave dataset) is more profound, due to advances in natural language processing, such as the Stanford NLP Parser, than the analysis used in Flesch-Kincaid [35], Coleman-Liau [10], and other readability formulas [31, 62].

3.6.3 Subject Headings and Reading Levels

Read_Level examines the subject headings[3] of a book *previously encountered* in books with a known readability level range recommended by their respective publishers. A previously encountered subject heading is a heading observed during the one-time mapping process, which paired subject headings assigned to each of the 9,000 books in the CLCD.com, a website established to assist teachers, parents, and librarians in choosing books for young readers, Young Adults Library Service Association (ala.org/yalsa), and Lexile.com dataset. To account for the possibility that a subject heading, SH, is paired with many books and thus many readability levels, Read_Level considers all readability levels paired with SH and computes the average of these levels as the reading level of SH. With the computed reading level of each subject heading, Read_ level considers the *average* of the reading levels of the subject headings assigned to a book Bk to predict the *reading level* of BK, since books that are *more difficult* to comprehend are often assigned *more* subject headings. We have empirically verified this claim by counting the number of subject headings assigned to each one of the 5,718 randomly chosen books (available at ARbookfind.com) with its readability levels determined by Accelerated Reader (AR) [49].

The mappings between the number of subject headings and grade levels are depicted in Figure 7. The trend line in Figure 7 has a positive slope of about $\frac{2}{9}$, which demonstrates that books of high readability levels are assigned, on average, more subject headings than books of lower reading levels.

[3] A *subject heading* is a set of *keywords* used by librarians to categorize and index books according to their themes. An example of a subject heading is "Fantasy—Mythical Creatures—Trolls—Green."

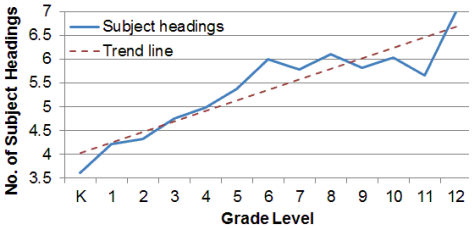


Fig. 7 The number of subject headings assigned to books versus their readability levels determined by AR

3.6.4 Book Authors and Readability Levels

Since most authors write for a particular group of readers with a certain grade level in mind [28], the *authors* of a book Bk are useful for partially determining the readability level of Bk. For example, a preview of a book by Agatha Christie, who writes adult crime novels, is harder to read than a preview of a book by Linda Aber, who is an author of children ghost stories. Given that the author of Bk is A, Read_Level assigns A to a particular reading level by (i) extracting the previews of books written by A, which serve as abstracts of books (co-)authored by A, (ii) determining the *subject heading* of each book of A and its corresponding reading level as discussed in Section 3.6.3, (iii) *averaging* the readability level of each book written by A that is determined by the subject headings. The resultant *averaged* reading level (rounded to the nearest whole number) determines the reading level A writes for, which is partially used to compute the reading level of Bk.

3.6.5 Assessing Readability Levels Using the Multi-Class SVM Model

To assign the readability level to a candidate book, Read_Level relies on SVM, a robust classifier [65]. As stated in [60], SVM is a machine learning methodology which has been shown to achieve high performance on classification (of readability levels in our case).

To train an SVM, it requires a set S which contains N (≥ 1) labeled training instances. In our case, training instances are of the form $\{(x_1, y_1), \ldots, (x_N, y_N)\}$, where x_i ($1 \leq i \leq N$) is an input vector of *features*, which are the features considered by Read_Level as introduced in Sections 3.6.1- 3.6.4, used for predicting the reading level of a candidate book i, and $y_i \in \{-1, 1\}$ is the corresponding class label of x_i, i.e., a reading level in Read_Level. Moreover, $\varphi(x_i)$ is the *kernel mapping* that generates the vector x_i in a feature space, and $K(x_i, x_j) = \varphi(x_i).\varphi(x_j)$ is the *kernel function* that determines the distance between $\varphi(x_i)$ and $\varphi(x_j)$ in the feature space. A *trained* SVM model is applied to determine the *reading level* of a book Bk predicted by Read_Level. In our implementation of SVM, we represent Bk as a vector v that consists of a value assigned to each *feature* exhibited in Bk, and determine the reading level that should be assigned to Bk, which is in the range of [12..19], i.e., from 12 to 19 of age.

Since the adapted SVM must handle more than two types of class labels, i.e., readability levels in our case, Read_Level considers the one-against-all [41] strategy in solving the multi-class prediction problem. Given j (> 2) different classes, the one-against-all approach constructs j binary SVM classifiers, each of which separates one class from the rest. The j^{th} SVM is trained using the training instances in which the instances belonged to the j^{th} class are given *positive* labels, and the remaining instances *negative* ones [41]. The one-against-one approach, on the other hand, constructs $\frac{j(j-1)}{2}$ binary classifiers, one for each possible class pair. In classifying a new instance, a vote is added to the class selected by each of the $\frac{j(j-1)}{2}$ binary classifiers and the class with the most votes is the one chosen as the class of the new instance. We adapt *one-against-all*, instead of one-against-one, in the implementation of SVM, since it is the most common multi-class categorization strategy [41]. The classification of a new instance v, i.e., the feature-vector representation of a given book, using the multi-class SVM is performed by selecting among the pre-defined reading levels (i.e., 12-19) the one for which the corresponding kernel function is the *highest*. We used 11,000 teenage books in the Book Cave dataset for training our multi-class SVM by partitioning the training dataset into 80/20, with 80% of them used for training the classifier and the remaining 20% for testing.

In implementing our multi-class SVM, we have adapted the Radial Basis Function (RBF) in the equation given below, which is one of the most typical kernels [66], as the *kernel function* K for the SVM.

$$K(x_i, x_j) = exp(-\frac{\|x_i - x_j\|^2}{\sigma^2}) \tag{6}$$

where $\|x_i - x_j\|$ is the Euclidean distance computed between vectors x_i and x_j[4], and σ is the parameter that determines the area of influence of the corresponding support vector. A large σ yields a smoother decision surface, since an RBF with a large σ allows a support vector to have a larger area of influence. σ is empirically determined to be 450 [66].

The smallest the difference between the predicted readability levels of a candidate book and a target book, the highest the feature value of the candidate book based on the readability level analysis.

3.7 CombMNZ

Based on the respective scores of the features, as discussed in Sections 3.2 through 3.6, computed for each candidate book of a target book, TBRec ranks all the candidate books accordingly. To compute a single score on which the cumulative effect of the five different features of each candidate book is determined for ranking propose, TBRec relies on the CombMNZ model, a well-established data fusion method for combining multiple ranked lists on an item I [11], i.e., a book in our case, to determine a *joint* ranking of I, a task known

[4]In Read_Level each vector, x_i and x_j, represents the heuristics, i.e., readability level features, of a book in a set of books.

as rank aggregation or data fusion.

$$CombMNZ_I = \sum_{c=1}^{N} I^c \times \mid I^c > 0 \mid, \text{ where } I^c = \frac{S^I - I^c_{min}}{I^c_{max} - I^c_{min}} \tag{7}$$

where N is the number of ranked lists to be fused, which is the *five* ranked lists in our case, I^c is the normalized score of I in the ranked list c, and $\mid I^c > 0 \mid$ is the number of non-zero, normalized scores of I in the lists to be fused. Prior to computing the ranking score of a candidate book CB, we transform the original scores in each feature ranked list of CB into a *common range* [0, 1] such that S^I is the score of I in the ranked list c to be normalized, I^c_{max} (I^c_{min}, respectively) is the maximum (minimum, respectively) score available in c, and I^c is the normalized score for I in c.

4 Experimental Results

Given a target book and a set of candidate books extracted using genres, we evaluate the performance of TBRec by (i) determining the *relevance* of its recommendations, (ii) deciding the *ranking accuracy* of its recommendations, and (iii) comparing its suggestions with three well-known book recommenders, Amazon, Barnes & Noble, and LibraryThing. The usefulness of a book recommended by either TBRec, Amazon, Barnes & Noble, or LibraryThing is determined by a number of independent appraisers, which serve as the *gold standard* of the conducted empirical study. Based on the ranking on relevant recommended books with respect to a given (i.e., target) book determined by individual appraisers, the *degree of accuracy* of each individual recommendation made by the four recommenders can be computed.

4.1 Emotion Analysis Using Book Descriptions

Recall that we create vector representations of the emotions expressed in the content of teenage books to estimate teenager's book preferences. We analyze emotions among adjacent age sets, or in other words, books that teenagers choose to read at a certain age. Using 11,000 books in the Book Cave dataset with each book labeled with an average age group, we notice a pattern of the emotion *joy* decreasing and the emotion of *sadness, fear, anger*, and *disgust* increasing as a teenager is getting older. This pattern becomes more evident with the analysis and it presents significant results ($p < 0.001$) in those emotion distributions between ages, especially between the age sets 12-13, 14-15 and 16-17. (See Figure 8 for the results.) These results are presented in more detailed in Table 4 with statistical significant results in those emotion distributions among different age groups **highlighted**. Each age group roughly corresponds to the ages of students of the respective grade level in high school in the USA.

Based on the results depicted in Figure 8, the emotion analysis on book content demonstrates an emotional shift preference among adjacent teenage groups. The results show that early teens incline towards the emotion *joy*, with

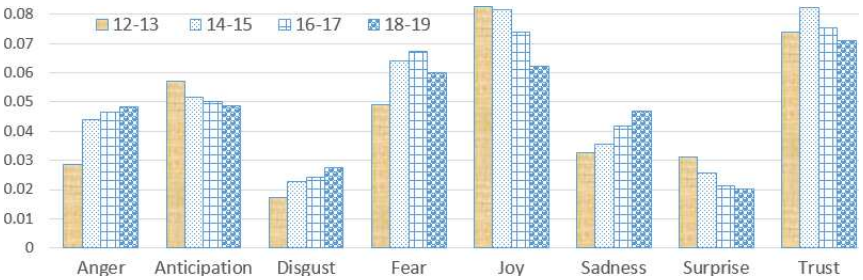


Fig. 8 Comparing emotional needs of teenagers using various teenage groups based on the Book Cave dataset

Table 4 Age group comparisons based on their emotion patterns with the statistical significance results highlighted

Emotion/Ages	12-13/ 14-15	12-13 16-17	12-13/ 18-19	14-15/ 16-17	14-15/ 18-19	16-17/ 18-19
Anger	**0.001**	**0.025**	**0.002**	0.213	0.678	0.334
Anticipation	0.138	0.727	0.301	**0.049**	**0.002**	0.128
Disgust	**0.005**	**0.001**	**0.001**	0.410	**0.026**	0.480
Fear	**0.002**	**0.001**	**0.002**	0.889	0.841	0.962
Joy	**0.006**	**0.012**	**0.013**	0.592	0.762	0.627
Sadness	**0.011**	**0.003**	**0.002**	0.372	0.168	0.659
Surprise	0.242	0.691	**0.038**	0.615	0.769	0.123
Trust	0.432	0.089	0.732	0.061	**0.045**	0.474

preference towards words such as celebration, excitement, gratitude, cheerful, and smile, which are words with high NRC-EIL intensity scores. This indicates that young teenagers prefer books with *happy* subjects as contrasted with books that tailor towards complex themes. As teenagers get older, the intensity of their feeling of great pleasure and happiness decreases, and their feeling of anxiety, safety, well-being, revulsion, or strong disapproval aroused by unpleasant or offense increase. They prefer words such as *bully, compassion, crime, battle*, and *mystery*, which are words with high NRC-EIL intensity scores. This indicates that older teenagers favor books with more elaborated themes such as *adventures*, *good* versus *evil*, and *romances*.

4.2 Data Source and Appraisers

To the best of our knowledge, there is no existing benchmark dataset that can be used for evaluating the performance of a teenage book recommender system. For this reason, we constructed our own dataset, using data made available for the public by BookCrossing [8], Goodreads [25], and Book Cave [7], three of the well-established websites for the book hobbyist community. BookCrossing, which is a free online book club, has established discussion forums, offered blogs, and accumulated more than 13.7 million books as of December 2021.

Table 5 Sources of data used by TBRec for feature analysis with data extracted using the API of the websites

Feature	Data Source	Dataset
Genres	Book Genres	Goodreads
Reader's Advisory	Book Reviews	Goodreads
Topic Analysis	Book Descriptions	Goodreads
Rating Prediction	User Ratings	BookCrossing
Emotion Traits	Book Content	Book Cave
Readability Level	Sample Content	Book Cave
Age groups (12-19)	53,000 Teenagers	BookCrossing

Goodreads, on the other hand, is a social cataloging website that provides the facility for each user to search its database of books and reviews, whereas Book Cave publishes a database of over 11,000 teenage books, and the web page for over 98% of books in the Book Cave database also includes a hyperlink to the Amazon Store web page for the Kindle (electronic) version of that book. From the Amazon Kindle Cloud Reader [56] web page we managed to collect the sample texts for the archived books. Table 5 shows the data sources used for the performance analysis of TBRec, Amazon, Barnes & Noble, and LibraryThing.

As there is no well-established standard nor existing book recommender system that can be adapted or used for comparing the performance of teenage book recommender systems, we turned to students at a local high school to conduct an empirical study that allows us to evaluate the performance of TBRec. We needed these students, who were in the 9^{th}- and 10^{th}-grade of their respective class at a local high school, to serve as the appraisers of our study, since they are the targeted users of TBRec and have read the target books used in the study so that we could collect their feedback on various recommendation tasks. A total of **56** students, who were 15- or 16-year-old, were recruited through their class teachers for our study using **12** target books, and the study was conducted between February 1 and March 31, 2022 at the high school. The fifty-six students and 12 target books are ideal numbers for our study according to the calculation based on the *cross-over experiment* [33] so that the experimental results are reliable and objective.

4.2.1 Determining the Number of Appraisers

In statistics, two types of errors, Types I and II, are defined [33]. Type I errors, also known as α errors or *false positives*, are the *mistakes* of *rejecting* a null hypothesis when it is true, whereas Type II errors, also known as β errors or *false negatives*, are the *mistakes* of *accepting* a null hypothesis when it is false. We apply the formula in [33] to determine the ideal number of appraisers, $n = 56$, which is dictated by the probabilities of occurrence of Types I and II errors, to evaluate minimizing the occurrence of Types I and II errors, to evaluate books suggested by the targeted book recommender systems.

$$n = \frac{(Z_{\frac{\alpha}{2}} + Z_{\beta})^2 \times 2\sigma^2}{\triangle^2} + \frac{(Z_{\frac{\alpha}{2}})^2}{2} \tag{8}$$

where $\triangle$ is the *minimal expected difference* to compare our recommendation approach with a librarian who manually chooses a book to be suggested to readers, which is set to 1 in our study as we expect our approach to make high-quality book recommendations as good as the ones made by librarians; σ^2 is the *variance*[5] of the recommended books, which is set to be 3.45 in our study (see the discussion on the computed value in the next paragraph); α (β, respectively) denotes the probability of making a Type I (II, respectively) error, which is set to be 0.05 (0.20, respectively), and 1 - β determines the probability of a false null hypothesis that is correctly rejected, and Z is the value assigned to the standard *normal distribution* of generated summaries.

Based on the standard normal distribution, when $\alpha = 0.05$, $Z_{\frac{\alpha}{2}} = 1.96$, and when $\beta = 0.20$, $Z_{\beta} = 0.84$. (See the explanations on setting the α and β values given below.) We conducted an experiment using a randomly sampled 50 target books to determine the value of σ^2. We chose only 50 books, since the *minimal expected difference* and *variance*, which are computed on a *simple random sample*, do not change with a larger sample set of books. σ^2 is computed by averaging the sum of the square difference between the mean and the actual number of *useful* recommendations[6] created for each one of the 50 target books. We obtained $\sigma^2 = 3.45$ for the recommendations.

The values of α and β are set to be 0.05 and 0.20, respectively, which imply that we have 95% *confidence* on the correctness of our analysis and that the *power* (i.e., probability of avoiding false negatives/positives) of our statistical study is 80%. According to [34], 0.05 is the commonly-used value for α, whereas 0.80 is a conventional value for 1 - β, and a test with $\beta = 0.20$ is considered to be statistically powerful. Based on the values assigned to the variables in Equation 8, The ideal number of appraisers used for our study is

$$n = \frac{(1.96 + 0.84)^2 \times 2 \times 3.45}{1^2} + \frac{1.96^2}{2} \cong 56 \tag{9}$$

4.2.2 Determining the Number of Target Books

To determine the ideal number of target books as test cases to be included in the controlled experiments, we rely on two different variables: (i) the *average attention span* of a teenager and (ii) the *average number of test cases* that a teenager can handle at a time. As mentioned in [58], the average attention span of a teenager is between twenty-five to thirty minutes to spend on a web search engine in one session. Based on this study, each appraiser was asked to evaluate the performance of each book recommender system involved in *three* test cases, i.e., target books, since evaluating all the *eight* recommendations[7] made for each one of the three test cases takes approximately 30 minutes, which falls into a teenager time span. Since each appraiser was committed to spend

[5] *Variance* is widely used in statistics, along with standard deviation (which is the square root of the variance), to measure the average dispersion of the scores in a distribution.

[6] A recommendation is considered *useful* if it is regarded as relevant to the corresponding target book determined by librarians recruited at a local school.

[7] Each recommendation is the snippet of the content of a book (limited to the first 500 characters) provided by the publisher of the book.

appropriately 120 minutes on the empirical study, we set up 4 sessions with 30 minutes each for each appraiser to conduct the empirical study. Altogether, a total of *12* ($= 4 \times 3$) target books were used for the study.

4.2.3 Our Verification approach

The students, who participated in the performance evaluation of TBRec (Amazon, Barnes & Nobles, and LibraryThing), were asked to determine which one of the eight recommendations[8], if there were any, were relevant books with respect to the corresponding target book. (We used the API provided by each of the three websites to harvest the book recommendations.) The two books marked as *relevant most often* by the appraisers were treated as the *gold standard* for the target book. Table 6 shows the top-2 recommendations for each one of the three sample target books suggested by TBRec, Amazon, Barnes & Noble, and LibraryThing, respectively, with the average ranking value of each recommended book made by the corresponding appraisers with respect to each target book.

4.3 Performance Evaluation

To evaluate the effectiveness of TBRec in recommending relevant and highly-ranked books to teenagers, we applied several performance measures commonly used in information retrieval and recommender systems [12]: *precision@1* ($P@1$), *precision@2* ($P@2$, since each recommender suggests two books), and *Mean Reciprocal Rank* (MRR), which measure the *top-ranked* and *first* useful recommendation among all the ranked recommendations, respectively.

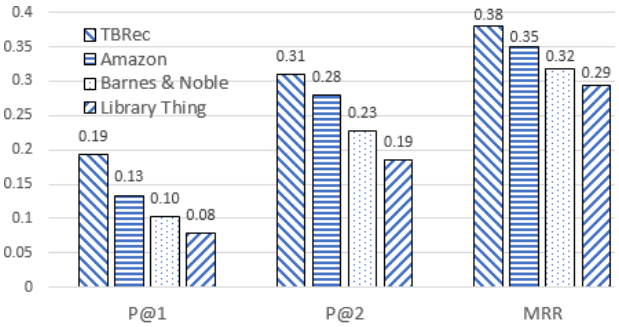
4.3.1 Accuracy of Filtering Candidate Books

In evaluating the performance of TBRec in terms of accurately filtering candidate books based on genres for feature analysis, we considered the books recommended by Amazon, Barnes & Noble, and LibraryThing based on the top-2 recommendations made by each book recommender according to the 12 target books. The genres of each target book form the *ground truth* of the empirical study, and we assume that each of the recommended books is relevant to the corresponding target book to a certain degree. The computed accuracy of the filtering approach based on book genres is **74%**, which indicates that using book genres every 3 out of 4 books are correctly filtered. The accuracy could *not* be higher likely because of the high degree of overlap between distinct genres assigned to different books, such as "High Fantasy" and "Epic Fantasy".

[8]Two each from TBRec, Amazon, Barnes & Nobles, and LibraryThing which were the top-2 recommendations made by the four recommender systems on a given target book, respectively. The appraisers had no idea which recommendation was made by which book recommender.

Table 6 Top-2 recommendations (*Recd*), for each of the three sample target books made by TBRec, Amazon, Barnes & Nobles, and LibraryThing and their respective average ranking value based on the gold standard established by the student appraisers

Target	**TBRec**		**Amazon**		**Barnes & Noble**		**LibraryThing**	
Book	1^{st} Recd	2^{nd} Recd	1^{st} Recd	2^{nd} Recd	1^{st} Recd	2^{nd} Recd	1^{st} Recd	2^{nd} Recd
The Golden Compass	The Yearling	The Lost City of Faar	The Two Towers	The Final Empire	His Dark Materials: Subtle Knife	The Amber Spyglass	Sabriel	Lirael
Avg. Rank	3.4	4.9	5.0	4.9	4.0	4.4	4.4	5.1
Of Mice and Men	North and South	The Stone Diaries	The Constitution of the USA	The War of the Worlds	The Grapes of Wrath	The Pearl	Lord of Files	East of Eden
Avg. Rank	3.4	4.8	4.3	4.9	4.5	3.8	5.6	5.1
White Teeth	Cut	Behind the Scenes at the Museum	Girl, Woman, Other	He Knew He was Right	The Flow	The Last One Out	On Beauty	Small Island
Avg. Rank	4.6	3.5	4.2	5.6	4.2	3.5	5.7	4.8
Hot Water Music	No Logo	The Last Man	Tales of Ordinary Madness	The Most Beautiful Woman in Town	Post Office	Ham on Rye	South of No North	The Road To LA Angeles
Avg. Rank	3.3	4.2	5.6	4.7	4.0	5.6	5.5	4.0
Overall Rank	**3.4**	**4.6**	**4.9**	**4.8**	**4.2**	**4.6**	**5.2**	**4.7**



(a) Comparing various book recommenders

(b) Comparing each individual feature and the combined features of TBRec

Fig. 9 Performance comparison of various book recommenders, in addition to each feature and combined features of TBRec

4.3.2 Comparing the Performance Evaluation of TBRec and Other Book Recommender Systems

Users of a recommender system tend to look at only the top few ranked results to find relevant recommendations. Some search tasks have only the top-ranked recommendation, i.e., $P@1$, in mind, whereas others might consider the top-3 ranked recommendations. After the gold standard for each one of the 12 test cases provided by the 56 appraisers were determined, we computed the $P@1$, $P@2$, and MRR values for various book recommenders involved in our empirical study. Figure 9(a) shows the performance metrics for the *average precision* at rank 1 (i.e., average $P@1$) and *average Precision* at rank 2 (i.e., average P@2), in addition to the average of the reciprocal ranks at which the *first* useful recommendation (among all the ranked recommendations) for each target book is made, i.e., MRR. The $P@1$, $P@2$, and MRR scores of TBRec are higher than the corresponding scores of Amazon, Barnes & Nobles, and LibraryThing, respectively, and the results are statistically significant based on the t-test ($p < 0.05$).

4.3.3 Feature Evaluation of TBRec

To evaluate the performance of each individual feature used by TBRec and its accuracy in making ranked recommendations on its own, we calculated their

$P@1$, $P@2$, and $nDCG$ values using one feature at a time to make recommendations. (nDCG [43], normalized discounted cumulative gain, penalizes useful suggestions ranked *lower* in the list of suggestions.) Hereafter, we calculated the same performance values using all the features combined, which is adapted by TBRec. All of these evaluations are based on the top-2 suggestions of each target book, and there are 20 test cases in this study. We have recruited 200 freshman students from our university, who are 18- or 19-year-old coming from various academic disciplines, to conduct the study. (Twenty and 200 were dictated by the cross-over experiment [33]). Given a target book, these students were asked to rank the relevance of each recommended book (based on a brief description of its content), and their evaluations yield the *ground truths* of our study.

According to the performance values shown in Figure 9(b), which uses different test cases compared with the ones used in Figure 9(a), we realize that the *Reader's Advisory* feature is the most promising one (with the exception of TBRec that uses all the features), which is followed by the *Emotion Traits*, since they perform significantly better than each of the other single features in terms of recommending and ranking relevant books to the users. Collectively, using all of the features, TBRec outperforms each individual feature, and the result is statistically significant based on the t-test ($p < 0.02$). The average $nDCG$, $P@1$, and $P@2$ values indicate that suggestions made by TBRec rank higher than the corresponding ranked books recommended by each individual feature as determined by the college students.

5 Conclusions

In the USA 73% of the population reads at least one book per year [22]. Book reading helps socialize the individual, especially teenagers [68]. In fact, books are an excellent medium to encourage teenage readers, since they are more likely to have a narrative than their virtual counterparts. In reading novels, teenagers can observe mature relationships outside their own [30]. As teenagers get older, they have their own flavors and tastes and are interested in choosing books on their own. Therefore, the book selection of a parent will be increasingly less relevant. Moreover, teenagers can see how different characters interact with each, and the results of the interactions. To help teenagers find suitable books more efficiently and easily, we have developed a unique book recommender, called *TBRec*. TBRec combines multiple book features to recommend books specific for teenagers. These book features, which include book genres, topic relevance, emotion traits, readers' advisory, predicted user rating, and readability level, have significant impact on the preference and satisfaction of teenagers on books. A conducted empirical study shows that by combining multiple book features to make book recommendations, TBRec performs significantly better than using only a single book feature. Moreover, books recommended by TBRec for teenagers are considered more favorable than the ones suggested by Amazon, Barnes & Nobles, and LibraryThing, respectively, the three well-known book recommenders.

References

[1] Ahmed B, Ghabayen A (2020) Review Rating Prediction Framework Using Deep Learning. Journal of Ambient Intelligence and Humanized Computing pp 1–10

[2] Alharthi H, Inkpen D, Szpakowicz S (2017) Unsupervised Topic Modelling in a Book Recommender System for New Users. In: Proceedings of the SIGIR 2017 eCom Workshop, p 8 Pages

[3] Alharthi H, Inkpen D, Szpakowicz S (2018) A Survey of Book Recommender Systems. Journal of Intelligent Information Systems 51(1):139–160

[4] Allington E, Gabriel E (2012) Every Child, Every Day. Educational Leadership 69(6):10–15

[5] Blei D, Ng A, Jordan M (2003) Latent Dirichlet Allocation. Journal of Machine Learning Research 3:993–1022

[6] Bobadilla J, Ortega F, Hernando A, et al (2012) A Collaborative Filtering Approach to Mitigate the New User Cold Start Problem. Knowledge-Based Systems 26:225–238

[7] Book Cave (2020) https://mybookcave.com/

[8] BookCrossing (2021) https://www.bookcrossing.com/

[9] Canada NRC (2020) https://saifmohammad.com/WebPages/AffectIntensity.htm

[10] Coleman M (1975) A Computer Readability Formula Designed for Machine Scoring. Applied Psychology 60(2):283–284

[11] Cormack G, Clarke C, Buettcher S (2009) Reciprocal Rank Fusion Outperforms Condorcet and Individual Rank Learning Methods. In: Proceedings of the International ACM SIGIR Conference on Research and Development in Information Retrieval. ACM, New York, pp 758–759

[12] Croft B, Metzler D, Strohman T (2010) Search Engines: Information Retrieval in Practice. Addison Wesley, San Francisco, California

[13] Currie D (1997) Decoding Femininity: Advertisements and Their Teenage Readers. Gender & Society 11(4):453–477

[14] Dali K (2014) From Book Appeal to Reading Appeal: Redefining the Concept of Appeal in Readers' Advisory. The Library Quarterly 84(1):22–48

[15] Davison A, Kantor R (1982) On the Failure of Readability Formulas to Define Readable Texts: A Case Study from Adaptations. Reading Research Quarterly 17(2):187–209

[16] Eldouma S, Adam S (2005) Relationship between Reading and Writing in English as a Second Language in the Context of Performance, Perceptions and Strategy Use. PhD thesis, Universiti Putra Malaysia

[17] Feldman R (2013) Techniques and Applications for Sentiment Analysis. Communications of the ACM (CACM) 56(4):82–89

[18] Ferrer E, Shaywitz B, Holahan J, et al (2015) Achievement Gap in Reading is Present As Early As First Grade and Persists through Adolescence. The Journal of Pediatrics 167(5):1121–1125

[19] Fry E (1968) A Readability Formula that Saves Time. Journal of Reading 11(7):513–578

[20] Garan E, DeVoogd G (2008) The Benefits of Sustained Silent Reading: Scientific Research and Common Sense Converge. The Reading Teacher 62:336–344. Https://doi.org/10.1598/RT.62.4.6

[21] Gelfand A (2000) Gibbs Sampling. American Statistical Association 95(452):1300–1304

[22] Gelles-Watnick R, Perrin A (2021) Who Doesn't Read Books in America?. https://www.pewresearch.org/fact-tank/2021/09/21/who-doesnt-read-books-in-america/

[23] Genre Literary Devices: Definition and Examples of Literary Terms. https://literarydevices.net/genre/

[24] Goodreads (2020) https://help.goodreads.com/s/article/How-do-I-get-a-copy-of-my-data-from-Goodreads

[25] Goodreads (2021) https://help.goodreads.com

[26] Gu Q, Zhou J, Ding C (2010) Collaborative Filtering: Weighted Nonnegative Matrix Factorization Incorporating User and Item Graphs. In: Proceedings of the 2010 SIAM International Conference on Data Mining (SDM), pp 199–210

[27] Guthrie J, Hoa A, Wigfield A, et al (2007) Reading Motivation and Reading Comprehension Growth in the Later Elementary Years. Contemporary Educational Psychology 32(3):282–313

[28] Hadaway N (2009) A Narrow Bridge to Academic Reading. Supporting English Language Learners 66(7):38–41

[29] Hill K (2013) The Arts and Individual Well-Being in Canada: Connections between Cultural Activities and Health, Volunteering, Satisfaction with Life, and Other Social Indicators in 2010. Hill Strategies Research Incorporated, Canada

[30] Howard V (2011) The Importance of Pleasure Reading in the Lives of Young Teens: Self-Identification, Self-Construction and Self-Awareness. Librarianship and Information Science 43(1):46–55

[31] Institute of School Renaissance. (2000) The ATOS Readability Formula for Books and How it Compares to Other Formulas. Tech. Rep. ED449468, ERIC Document Reproduction Service

[32] Iyengar S, Lepper M (2000) When Choice is Demotivating: Can One Desire too much of a Good Thing? Personality and Social Psychology 79(6):995

[33] Jones B, Kenward M (2003) Design and Analysis of Cross-Over Trials, 2nd Edition. Chapman and Hall

[34] Kazmier L (2003) Schaum's Outline of Business Statistics. McGraw-Hill

[35] Kincaid J, Fishburne R, Rogers R, et al (1975) Derivation of New Readability Formulas (Automated Readability Index, Fog Count, and Flesch Reading Ease Formula) for Navy Enlisted Personnel. Tech. Rep. 8-75, Chief of Naval Technical Training

[36] Kingma D, Welling M (2013) Auto-Encoding Variational Bayes. arXiv preprint arXiv:1312.6114

[37] Koren Y, Bell R (2015) Advances in Collaborative Filtering. Recommender systems handbook pp 77–118

[38] Kowalczyk P (2021) There are Now Over 10 Million Publications in the Kindle Store. https://ebookfriendly.com/over-10-million-kindle-ebooks-on-amazon/

[39] Lee J (1997) Analyses of Multiple Evidence Combination. In: Proceedings of the International ACM SIGIR Conference on Research and Development in Information Retrieval. ACM, New York, pp 267–276

[40] Liang Z, Huang S, Huang X, et al (2020) Post-Click Behaviors Enhanced Recommendation System. In: Proceedings of the IEEE 21^{st} International Conference on Information Reuse and Integration for Data Science (IRI), pp 128–135

[41] Liu Y, Zheng Y (2005) One-Against-All Multi-Class SVM Classification Using Reliability Measures. In: Proceedings of International Joint Conference on Neural Networks (IJCNN'05). IEEE, pp 849–854

[42] Love K, Hamston J (2004) Committed and Reluctant Male Teenage Readers: Beyond Bedtime Stories. Journal of Literacy Research 36(3):335–400

[43] Manning C, Raghavan P, Schutze H (2008) Introduction to Information Retrieval. Cambridge University Press, New York

[44] Milton A, Green M, Keener A, et al (2019) StoryTime: Eliciting Preferences from Children for Book Recommendations. In: Proceedings of the 13^{th} ACM Conference on Recommender Systems (RecSys). ACM, New York, pp 544–545

[45] Milton A, Batista L, Allen G, et al (2020) "Don't Judge a Book by Its Cover": Exploring Book Traits Children Favor. In: Proceedings of the 14^{th} ACM Conference on Recommender Systems (RecSys). ACM, New York, pp 669–674

[46] Minka T (2013) Expectation Propagation for Approximate Bayesian Inference. arXiv preprint arXiv:1301.2294

[47] Mohammad S (2012) From Once Upon a Time to Happily Ever After: Tracking Emotions in Mail and Books. Decision Support Systems 53(4):730–741

[48] Mortiboys A (2013) Teaching with Emotional Intelligence: A Step-by-Step Guide for Higher and Further Education Professionals. Routledge

[49] Pavonetti L, Brimmer K, Cipielewski J (2002) Accelerated Reader: What are the Lasting Effects on the Reading Habits of Middle School Students Exposed to Accelerated Reader in Elementary grades? Adolescent and Adult Literacy 46(4):300–311

[50] Pera M (2009) Improving Library Searches Using Word-Correlation Factors and Folksonomies. Master's thesis, Brigham Young University, Provo, Utah

[51] Pera M (2014) Using Online Data Sources to Make Recommendations on Reading Material for K-12 and Advanced Readers. PhD thesis, BYU

[52] Plutchik R (1997) Circumplex Models of Personality and Emotions, American Psychological Association, Washington, D.C., chap The Circumplex as a General Model of the Structure of Emotions and Personality, pp 17–45

[53] Porteous I, Newman D, Ihler A, et al (2008) Fast Collapsed Gibbs Sampling for Latent Dirichlet Allocation. In: Proceedings of ACM SIGKDD Conference on Knowledge Discovery and Data Mining. ACM, New York, pp 569–577

[54] Putri T, Zulkarnain (2020) Proposed Model of Academic Reading Material Recommendation System. In: Proceedings of the 3^{rd} Asia Pacific Conference on Research in Industrial and Systems Engineering (APCORISE 2020), pp 105–109

[55] Rapport N, Dawson A (2021) The Topic and the Book. In: Migrants of Identity. Routledge, p 3–17

[56] Reader AKC (2021) https://read.amazon.com/

[57] Reagan A, Mitchell L, Kiley D, et al (2016) The Emotional Arcs of Stories are Dominated by Six Basic Shapes. EPJ Data Science 5(1):1–12

[58] Rozakis L (2002) Test Taking Strategies and Study Skills for the Utterly Confused. McGraw Hill

[59] Saricks J (2005) Readers' Advisory Service in the Public Library, 3^{rd} Ed. ALA American Library Association Store, Atlanta, GA

[60] Sculley D, Wachman G (2007) Relaxed Online SVMs for Spam Filtering. In: Proceedings of the International ACM SIGIR Conference on Research and Development in Information Retrieval. ACM, New York, pp 415–422

[61] Shu J, Shen X, Liu H, et al (2018) A Content-Based Recommendation Algorithm for Learning Resources. Multimedia Systems 24:163–173

[62] Smith D, Stenner A, Horabin I, et al (1989) The Lexile Scale in Theory and Practice: Final Report. Tech. Rep. ED307577, ERIC Document Reproduction Service

[63] Spache G (1953) A New Readability Formula for Primary-Grade Reading Materials. Elementary School 53(7):410–413

[64] Strommen L, Mates B (2004) Learning to Love Reading: Interviews with Older Children and Teens. Adolescent & Adult Literacy 48(3):188–200

[65] Suthaharan S (2016) Support Vector Machine. In: Machine Learning Models and Algorithms for Big Data Classification. Springer, p 207–235

[66] Tang B, Mazzoni D (2006) Multiclass Reduced-set Support Vector Machines. In: Proceedings of the 23^{rd} International Conference on Machine Learning (ICML). ACM, New York, pp 921–928

[67] Taylor J, Hora A, Krueger K (2019) Self-Selecting Books in a Children's Fiction Collection Arranged by Genre. Journal of Librarianship and Information Science 51(3):852–865

[68] Tveit A, Mangen A (2014) A Joker in the Class: Teenage Readers' Attitudes and Preferences to Reading on Different Devices. Library & Information Science Research 36(3-4):179–184

[69] Wanzek J, Vaughn S, Kim A, et al (2006) The Effects of Reading Interventions on Social Outcomes for Elementary Students with Reading Difficulties: A Synthesis. Reading & Writing Quarterly 22(2):121–138

[70] Wikipedia (2021) https://en.wikipedia.org/wiki/Wikipedia:Database_download

[71] WordNet (2021) https://wordnet.princeton.edu

[72] Xin Y, Chen Y, Jin L, et al (2017) TeenRead: An Adolescents Reading Recommendation System Towards Online Bibliotherapy. In: Proceedings of the 2017 IEEE International Congress on Big Data (BigData Congress), pp 431–434

[73] Zhou Y (2020) Design and Implementation of Book Recommendation Management System based on Improved Apriori Algorithm. Intelligent Information Management 12:75–87

[74] Zuo L, Xiong S, Qi X, et al (2021) Communication-Based Book Recommendation in Computational Social Systems. Complexity 2021:10 Pages