Assisting Emergent Readers in Finding Books to Read

Maria Soledad Pera
Department of Computer Science
Boise State University
Boise, Idaho, U.S.A.
soledadpera@gmail.com

Yiu-Kai Ng
Computer Science Department
Brigham Young University
Provo, Utah, U.S.A.
ng@compsci.byu.edu

ABSTRACT

It is imperative to motivate children to read by offering them appealing books so that they can gradually establish a reading habit during their formative years. However, with the huge volume of existing and newly-published books, it is a challenge to find the right ones that match children's interests and readability levels. In response to the needs, we have developed K3Rec, an unsupervised recommender, which suggests books that match the *interests/preferences* and *reading abilities* of emergent (i.e., K-3) readers.

Categories and Subject Descriptors

H.3.3 [**Information Search and Retrieval**]: Information Filtering, Retrieval Models, Selection Process

Keywords

Book recommendation system, emergent readers, K-3

1. INTRODUCTION

Learning to read is a key milestone for children, specially given that reading provides the foundation for children's academic success. In fact, children who "do not read proficiently by the end of third grade are four times more likely to leave school without a diploma than proficient readers."[1] The aforementioned findings constitute the essence of encouraging good reading habits early on. Identifying books appealing to emergent readers,[2] however, can be challenging, given the amount of available books that address a diversity of topics and target readers at various reading levels.

In the quest for locating books which can help improve the reading skills of K-3 readers, parents/educators/children can turn to online book recommenders. Unfortunately, these recommenders require user-defined information, such as ratings and accessing patterns, to make suggestions for respective users. Personal information of K-3 users, however, may not exist due to the lack of online networking sites targeting K-3 users or may not be publicly accessible due to the ethical obligation to respect the online privacy of children. Moreover, majority of these recommenders fail to explicitly consider (i) the reading ability of a reader, and/or (ii) unique characteristics that distinguish books targeting emergent readers. To address these issues, we have developed *K3Rec*, an unsupervised book recommender specially designed to suggest books for K-3 readers, an audience who has not been catered by existing recommenders. Unlike existing state-of-the-art book recommenders, K3Rec does not rely on the availability of *user-defined information* to make book suggestions. Instead, K3Rec takes advantage of book metadata, which are either readily accessible or can be inferred from reputable online data sources, such as book reviews, that are publicly available from book-related websites. K3Rec is unique, since it explicitly considers the *illustrations* of books for emergent readers.

2. OUR PROPOSED RECOMMENDER

In making recommendations for a K-3 reader R, K3Rec first analyzes a given book B known to be of interest to R and identifies books that are compatible with the readability level of R.[3] These books are treated as *candidate books* to be considered for recommendation and are selected among the books available at a book repository, such as (i) *reputable websites*, e.g., OpenLibrary.org, (ii) *school/public libraries*, and (iii) book-related *bookmarking sites*, e.g., BiblioNasium.com. K3Rec analyzes each candidate book CB based on diverse publicly accessible book metadata (see details below). Hereafter, K3Rec computes a single ranking score of CB using CombMNZ and presents the top-ranked candidate books as suitable suggestions for R.

Content Analysis. K3Rec computes the *content similarity* score between CB and B by analyzing the bag-of-(nonstop, stemmed) word representation of the descriptions of CB and B, which can be extracted from book-related websites, such as Amazon.com and Library of Congress (catalog.loc.gov), and by using Word-correlation factors [2]. K3Rec prioritizes candidate books based on their degrees of similarity with B.

---

[1] http://goo.gl/HQrPOA

[2] Emergent (or early) reading refers to the knowledge, skills, and dispositions acquired in reading (and writing) in primary school grades prior to and up till the 3rd grade [4].

---

[3] *Compatible books* are books within half a grade of the grade level of a book B (as determined using TRoLL [2]) that is of interest to a reader R. This measure is ensured that the compatible books are appropriate for R.

Illustration-Based Analysis. Since illustrations play a role in "directly encouraging children's emergent literacy development" [1], K3Rec considers book illustrations as part of its recommendation process. While book illustrations are not always freely accessible due to copyright laws, there are a number of websites that offer access to book covers (e.g., Google Books[4]). K3Rec takes advantage of such resources and calculates a score that reflects the *degree of resemblance* between the book covers of CB and B. It is not an easy task, however, to compute the aforementioned score, given that the similarity between images is based on identifying same/similar object(s) or scene(s) even if they are presented under different conditions, such as viewpoint changes, image blur, and illumination changes. K3Rec applies OpenCV, an open source computer vision/machine learning library to determine the similarity between any two book covers.

Topical Analysis. K3Rec examines *topical information* of CB to determine its suitability for R based on Library of Congress Subject Headings (LCSH) assigned to CB. K3Rec considers the *count* of LCSH assigned to CB, since books that are *more difficult* to comprehend are often assigned *more* LCSH [2]. In addition, K3Rec considers the grade levels associated with LCSH assigned to CB and determines the proportion of LCSH of CB that are associated with grade levels similar to the grade level of R (through book B).

Book-Length Analysis. As stated in [3], books for emergent readers are on an average of 32 pages in length. Relatively short books are preferred, since they can be read in one (or very few) sittings, which offers their readers a sense of accomplishment in finishing a book. K3Rec measures the degree to which the length of CB is within 32 pages and imposes a penalization on books longer than 32 pages.

Writing Style-Based Analysis. Another characteristic often applied to books for emergent readers is the *simplicity* and *directness* of their texts. Identifying the writing style of books, however, is non-trivial for the lack of access (due to copyright laws) to sample text on books required to perform semantic/syntactic analysis. For this reason, K3Rec relies on ABET [2] to obtain a description of the writing style of each candidate book CB. Using the ABET-generated *writing style description* of CB, K3Rec quantifies the *degree of directness* and *simplicity* of (the textual content of) CB.

Rating Assessment. As product ratings capture an independent measure of the quality of a product based on the opinions of appraisers, it is natural for K3Rec to prioritize books that have been assigned a high rating (on Google Books or similar book-related websites). Note that unlike existing book recommenders [5], K3Rec does not rely on the availability of personal ratings assigned to books by an individual user (to reflect the degree to which a book matches his interests), which are seldom made by K-3 readers.

3. EXPERIMENTAL RESULTS

We have conducted two empirical studies to assess the performance of K3Rec: one relies on data from BiblioNasium.com, which includes 1,705 K-3 users and their bookmarks, and the other on data collected using Amazon's Mechanical Turk.[5] In either study, K3Rec uses close to 20,000 books available at BiblioNasium.com as its book repository.

Using data from BiblioNasium, we verified that K3Rec outperforms ($p < 0.001$) BReK12 [2] in terms of *Normalized Discounted Cumulative Gain*. We compared K3Rec with BReK12, since to the best of our knowledge BReK12 is the only recommender that explicitly considers the readability level of its users in making book recommendations.

To evaluate the degree to which books recommended by K3Rec are preferred over those suggested by recommendation modules at book-related websites, we first selected recommenders that adopt diverse strategies in making suggestions: (i) Amazon, which considers purchasing patterns of its users, (ii) GoodReads,[6] which "combines multiple proprietary algorithms that analyze 20 billion data points", and (iii) NoveList,[7] which examines a number of book-related information, including title and appeal factors. Hereafter, we required Mechanical Turk users to select among a set of possible choices (generated using the aforementioned recommenders and K3Rec) the top-2 recommendations most closely related to each test book B, which were treated as the *gold standard* for B. Based on the 400 responses collected during the month of April 2014, we computed the accuracy of the top-2 recommendations made by K3Rec and each of the recommenders considered for comparison purpose. Recommendations made by K3Rec are preferred over those made by Amazon, GoodReads, and NoveList ($p < 0.05$ and $p < 0.001$, respectively). Moreover, in making recommendations for emergent readers, K3Rec considers books provided directly by K-3 readers (or their parents/teachers) to generate suggestions. Recommendations made by Amazon that target children, however, are the results of extensive analysis of the purchasing patterns of adults, which might not accurately reflect the interests of emergent readers in books.

4. CONCLUSIONS

We have introduced K3Rec, an unsupervised book recommender developed for K-3 readers who are not currently targeted by existing recommenders. Unlike current recommenders, K3Rec does not rely on personal data, such as ratings or bookmarks, which are rarely created by emergent readers, to make recommendations. Conducted experiments using data from BiblioNasium and a crowdsourcing platform have verified the relevance of books suggested by K3Rec.

5. REFERENCES

[1] L. Justic and J. Kaderavek. Using Shared Storybook Reading to Promote Emergent Literacy. *Teaching Exceptional Children*, 34(4):8–13, 2002.

[2] M. Pera. *Using Online Data Sources to Make Recommendations on Reading Materials for K-12 and Advanced Readers*. PhD Thesis, BYU, April 2014.

[3] M. Renck. *Young Children and Picture Books (2nd Ed.)*. National Association for the Education of Young Children, 2004.

[4] K. Roskos, J. Christie, and D. Richgels. The Essentials of Early Literacy Instruction. *Young Children*, 58(2):52–60, 2003.

[5] C. Yang, B. Wei, J. Wu, Y. Zhang, and L. Zhang. CARES: A Ranking-oriented CADAL Recommender System. In *ACM/IEEE JCDL*, pages 203–212, 2009.

---

[4]http://books.google.com
[5]https://www.mturk.com/mturk/welcome

[6]goo.gl/99me5f
[7]support.epnet.com/knowledge_base/detail.php?id=4772