TeenRead: An Adolescents Reading Recommendation System Towards Online Bibliotherapy

Yunxing Xin, Yongqiang Chen, Li Jin, Yici Cai, Ling Feng
Dept. of Computer Science and Technology
Centre for Computational Mental Healthcare Research, Institute of Data Science, Tsinghua University, Beijing, China
Email: xinyunxing0001@163.com, l-jin12@mails.tsinghua.edu.cn, {chenyongqiang,caiyc,fengling}@tsinghua.edu.cn

***Abstract*—Bibliotherapy has been proved to be an effective way to deal with adolescents psychological stress. Specific reading materials are provided to patients with physical or mental diseases for the purpose of prevention, healing, and rehabilitation. But traditional bibliotherapy requires professional staff with the background of both psychological and library services, which is quite demanding and labor consuming. Moreover, bibliotherapy based on paper books is getting ill-fitted in the present big data era. To address the limitations, this paper proposes an online reading recommendation system called *TeenRead* to carry out bibliotherapy for adolescents. *TeenRead* involves the management of users and articles, analysis of users' dynamic reading behaviors, as well as the recommendation based on users' stress categories, stress levels, and reading interests. The results of the user study on 10 volunteers show that, the average decrease of users' stress level is significantly dropped by 22% after a period of reading on *TeenRead*, which proves that *TeenRead* performs pretty well as a new method of bibliotherapy.**

***Keywords*-Bibliotherapy; adolescent; stress easing; reading interest; reading recommendation.**

I. Introduction

Nowadays, adolescents are bearing different psychological stress from academic research, communication, family, emotion, self-awareness, employment, and so on. Many psychological methods have been proposed to help coping with adolescent stress, such as mindfulness meditation, decompression Yoga, sports, and mental health care education. But most of the decompression methods require either people's full attention, which is quite time consuming, or more or less professional cares, which are difficult to be carried out by adolescents. Considering the real condition of adolescents, bibliotherapy (also called as *reading therapy*), which serves as a psychological treatment by reading some specific books or related information under the scientific guidance [1], could be an effective way to ease stress. However, bibliotherapy also faces many difficulties. First of all, it is demanding and labor consuming to make bibliotherapy in practice. As the person, who is responsible for bibliotherapy, is required of the knowledge of not only psychology but also library reader services [2]. Besides, for every stressful person, there must be a certain amount of reading materials customized for her/him. Secondly, the traditional methods of bibliotherapy may be unsuitable in the current reading environment any more. Nowadays, adolescents are more likely to read through WWW and mobile phones other than paper books.

To deal with these problems of bibliotherapy, this paper builds a reading recommendation system called *TeenRead*, which automatically recommends reading materials to adolescents based on their stress and reading interests. *TeenRead* serves as a web application so that users can access it through any platform with a web browser. The major contributions of this work are two-folds. Firstly, We discover a new method of bibliotherapy, which combines reading recommendation and stress easing in the age of information and big data era. Secondly, We find a way to capture users' implicit reading behaviors and gradually tune *TeenRead*'s recommendation to tailor to users' interests and stress. Our user study shows *TeenRead* could be a convenient and effective means of bibliotherapy to assist teens to release stress.

The remainder of the paper is organized as follows. Section 2 gives a brief introduction to bibliotherapy and some related work on computer based bibliotherapy and recommendation. Section 3 and 4 detail the key reading recommendation module and users reading behavior analysis module, respectively. We conduct an user study to evaluate the effectiveness of *TeenRead* in Section 5. Finally, we give a conclusion in Section 6.

II. Related Work

Bibliotherapy is recognized as an effective method of psychotherapy, and the research on bibliotherapy has lasted hundreds of years in the west. In 1848, John M. Galt read the paper in the annual convention of the American Psychological Association (APA) about the effects of bibliotherapy, and analyzed the classification of patients and corresponding reading measures. This paper is considered to be the first one in bibliotherapy [3]. The formal and systematic research on bibliotherapy began with Samuel Mc Chord Crothers [4]. Nowadays, bibliotherapy has formed a fairly complete theory system and become an important research content in both library science and medical rehabilitation.

In the era of information, computer has been widely involved in bibliotherapy. Peli-Hasz et al. [5] expand the definition of bibliotherapy and take e-bibliotherapy (computer-

978-1-5386-1996-4/17 $31.00 © 2017 IEEE
DOI 10.1109/BigDataCongress.2017.63

/Internet-based bibliotherapy) as an strong trend of bibliotherapy. In real practice, Wilawan [6] developed an e-bibliotherapy system which allows youth to manually share contents and build friendship network.

Different from the previous work, this study aims to carry out online bibliotherapy towards adolescents by automatically reading recommendation. In order to make *TeenRead* effective in easing users' stress, the recommended articles should not only be specific to user's stress categories, but also be attractive to users to make them willing to read. Recommendation approaches are comprehensively compared in our study. Beel et al. generally divide them into seven categories, namely, *stereotyping, content-based filtering, collaborative filtering, co-occurrence, graph-based, global relevance,* and *hybrid* [7]. Considering the pertinence of the recommended articles for bibliotherapy, our recommendation is based on CBF. In addition, we also propose several ways to deal with the cold-start problem, to balance the stress easing effects and interests of recommended articles, and to update recommendation based on users' dynamic reading behaviors.

III. Reading Recommendation

A. Guidelines for Reading Recommendation

1) Stress Easing by Reading: The primary goal of *TeenRead* is to alleviate teens psychological stress by reading. Therefore, the recommendation of articles mainly focuses on the effect of easing stress. According to psychological research, six categories of stress that teens usually encounter are considered, i.e., SC={*study, family, affection, inter-personnel, self-recognition, employment*}, and $|SC|$=6. Stress levels take values from 0 to 5, SL={0,1,2,3,4,5}, denoting *no stress* to *very strong stress*

2) Reading Interests: On the other hand, users reading interests are also a key factor for consideration. If the recommended articles are not attractive to the users, they won't be read, and thus make no attributions to the stress release. Ideally, the articles with the best effect of stress easing are the same as those with the highest interest degree for the users. Nevertheless, in reality, it's difficult to reach the maximum on both aspects, so it is necessary and challenging for *TeenRead* to make a tradeoff between these two.

3) Unification of Stress Easing and Reading Interest: To unify the two factors (stress easing and reading interest) into recommendation, we classify each article into different categories corresponding to stress categories. Due to the diversity and huge volume of articles, a finer grained categorization is performed, as reflected from the following two perspectives.

- Article's category set (denoted as AC) encloses stress category set SC, where $SC \subset AC$. In *TeenRead*, AC={*study, family, affection, inter-personnel, self-recognition, employment, life, art, humor, book*}.
- An article categorization is further divided into a number of sub-categories. Take *study* for example, its sub-categories are {*university, course, resource, competition, rank, dual-degree, final year project, exempt from postgraduate, recommendation, postgraduate entrance exam, study-abroad*}. Let ACs denote the set of all sub-categories, and $|ACs|$ denotes the total number of articles sub-categories.

An article may belong to multiple categories and multiple sub-categories. For example, an article entitled "*Advice into Society*" gives suggestions on job, personnel relation, self-cognition, etc., falling into multiple categories and sub-categories.

B. Method of Reading Recommendation

1) Notions of Stress Easing Effect Vector and Interest Vector: Users may experience different types of stress. For each user, *TeenRead* computes and maintains *a stress easing effect vector* $\overline{E}$, together with *an interest vector* $\overline{I}$. The former records the stress easing effects by reading different categories of articles, while the later records the interests of reading different categories and sub-categories of articles.

Definition 1: Let SC be the set of stress categories. Corresponding to the six typical stress categories, an **user's stress easing effect vector** is defined as $\overline{E}$=$(c_1.ease, c_2.ease, \cdots, c_{|SC|}.ease)$, where for $\forall i$ $(1 \leq i \leq |SC|) \wedge (c_i \in SC) \wedge (c_i.ease \in [0,5])$. Initially, $c_1.ease$=$\cdots$=$c_{|SC|}.ease$=0. □

Definition 2: Let ACs be the set of articles' sub-categories. An **user's interest vector** is defined as $\overline{I} = (cs_1.int, cs_2.int, \cdots, cs_{|ACs|}.int)$, where for $\forall i$ $(1 \leq i \leq |ACs|) \wedge (cs_i \in ACs) \wedge (cs_i.int \in [0,1])$.

Initially, $cs_1.int$=$\cdots$=$cs_{|ACs|}.int$=0. □

2) Recommendation based on Stress Easing Effect Vector and Interest Vector: For simplicity, assume $SC(article)$ and $ACs(article)$ return the set of categories in SC and set of sub-categories in ACs which $article$ falls into.

We measure an user's interest in $article$ by its RSS (root of squares sum) of the interest values of article's sub-categories.

$$Interest(\overline{I}, article) = \sqrt{\sum_{cs \in ACs(article)} cs.int^2}$$

Also, $article$'s stress easing effect is measured by its RSS of categories's stress easing effects.

$$Ease(\overline{E}, article) = \sqrt{\sum_{c \in SC(article)} c.ease^2}$$

Taking both interest and stress easing into account, the computation of an article 's recommendation score can be defined as follows.

Definition 3: Given an user's stress easing effect vector $\overline{E}$ and an interest vector $\overline{I}$. the **recommendation score of** $article$ **to the user** is defined as:

$$Recommend(\overline{E}, \overline{I}, article) = \rho * Ease(\overline{E}, article) + (1-\rho) * Interest(\overline{I}, article)$$

where parameter $\rho \in [0, 1]$ is a coefficient for adjusting the weights of stress easing effect and reading interest. In the study, ρ=0.5. □

Note that at the initial start-up phase when $\overline{E}$=(0, $\cdots$, 0) and $\overline{I}$=(0, 0, $\cdots$, 0), all articles's recommendation scores are 0. In this case, *TeenRead* will randomly select articles to recommend. While user's reading, these two vectors are updated by **users reading behavior analysis module**.

IV. Users Reading Behavior Analysis

This module is responsible for dynamically tuning user's interest vector and stress easing effect vector to tailor to users' real interests and stress status.

A. Maintenance of User's Interest Vector $\overline{I}$

$\overline{I} = (cs_1.int, cs_2.int, \cdots, cs_{|ACs|}.int)$ reflects user's interested (sub-)categories. Each element cs_i (where $1 \leq i \leq |ACs|$) of $\overline{I}$ is a sub-category in ASc, and $cs_i.int \in [0, 1]$ is the interest degree of the user in sub-category cs_i, where cs_i=0 indicates no interest at all, and cs_i=1 indicates a complete interest. Initially, $\overline{I}$=$(0, 0, \cdots, 0)$.

Maintenance of user's interest vector $\overline{I}$ is conducted under the following two situations.

1) Re-Setting Up $\overline{I}$ Upon User's Explicit Selection: When the user explicitly selects/updates the interested (sub-)categories anytime in his/her personal profile, *TeenRead* will firstly initialize $\overline{I}$=$(0, 0, \cdots, 0)$. For the non-selected sub-categories, their interest degrees are set to 0.

Beyond that, considering the content similarity among sibling sub-categories under the same category (e.g., *post-graduate study* and *study-abroad*), if an user is interested in cs_i, s/he may probably like other similar sibling sub-categories. Therefore, followed by each user's manual selection, we increase those non-selected sub-categories' interest degrees based on their similarity with the selected sibling sub-categories. Let $cs_i, cs_k \in ACs$ be two sibling sub-categories under the same category. Their content similarity is measured by function $Sim(cs_i, cs_k) \in [0, 1]$. Assume cs_i is not manually selected by the user as an interested sub-category, but its sibling cs_k is selected, we refine $cs_i.int$=$max_{cs_k} Sim(cs_i, cs_k)$.

In the current study, $Sim(cs_j, cs_k)$=0.2 for any two sibling sub-categories. After re-setting up all the elements of $\overline{I}$, we normalize the length of $\overline{I}$ to be 1: $\overline{I} = \frac{\overline{I}}{||\overline{I}||_2}$.

2) Adjusting $\overline{I}$ Based On User's Implicit Reading Behaviors: If the user does not explicitly select interested sub-categories, *TeenRead* dynamically updates $\overline{I}$ every w days according to user's reading behaviors during the past w days. User's interest in a sub-category is updated based on the following three factors:

(1) reading priority f_{seq} of articles containing this sub-category among the m recommended articles. Reading priority is an important measurement of user's interest. For simplicity of description, let $seq(article_{cs_i,k})$ indicates the reading order of the kth article which belongs to the ith sub-category cs_i, where $k \in [1, m]$, and $cs_i \in ACs$. If $article_{cs_i,k}$ is not read, we set $seq(article_{cs_i,k})$ to $m+1$. We assume that, the earlier $article_{cs_i,k}$ is read, the larger its reading priority should be. So $f_{seq}(article_{cs_i,k})$ can be defined as follows:

$$f_{seq}(article_{cs_i,k}) = \frac{(m + 1 - seq(article_{cs_i,k}))}{m}$$

Bacause one article may belongs to many sub-categories, we make the maximum value of $f_{seq}(article_{cs_i,k})$ to be the real f_{seq} value of sub-category cs_i, which is:

$$f_{seq}(cs_i) = max_{k \in [1,m]} f_{seq}(article_{cs_i,k})$$

(2) whether the user has supported, commented, collected and shared articles containing this sub-category or not f_{act}. We take the above four actions into account for the reason that they are positively related to user's interests. We use four vectors SupportVec, CommentVec, CollectVec and ShareVec to store the frequency of the mentioned four reading actions to articles belong to the sub-category during the past w days. Taking SupportVec for example, the element of it is $support_{cs_i}$, indicating his/her supporting number to articles belonging to cs_i. So $support_{cs_i}$ is calculated like:

$$support_{cs_i} = \sum_{k \in [1,m]} doSupport(article_{cs_i}, k)$$

Where $doSupport(article_{cs_i,k}) = 1$ if the user have supported $article_{cs_i,k}$; otherwise it's 0;

The calculation of elements in CommentVec, CollectVec and ShareVec is similar to that of SupportVec. Let $comment_{cs_i,k}$, $collect_{cs_i,k}$ and $share_{cs_i,k}$ respectively denote the element of each vector above. To combine the effects of these four vectors, we standardize each elements to be $0 \sim 1$, so f_{act} can be calculated as:

$$f_{act}(cs_i) = (\frac{support_{cs_i}}{Max(SupportVec)} + \frac{comment_{cs_i}}{Max(CommentVec)} + \frac{collect_{cs_i}}{Max(CollectVec)} + \frac{share_{cs_i}}{Max(ShareVec)})/4$$

(3) time spent on reading articles containing this sub-category f_{time}. The longer the time spent on $article_{cs_i,k}$ is, the more the user is interested in sub-category cs_i. Even though this assumption may not be valid in all situations, it can be applied to most situations. In addition, the time spent on reading is also influenced by the length of the article. So we take the ratio of time spent on reading to the length of this article to illustrate f_{time}, whose value for the past w days can be calculated as

$$f_{time}(int_i) = \sum_{k \in [1,m]} \frac{ReadTime(article_{cs_i,k})}{Len(article_{cs_i,k})}$$

where $ReadTime(article_{cs_i,k})$ is the time spent on reading $article_{cs_i,k}$, and $Len(article_{cs_i,k})$ is the length of $article_{cs_i,k}$ represented by the number of characteristics it.

After the values of f_{seq}, f_{act} and f_{time} are obtained, we can update $\overline{I}$ with their comprehensive effects. Let $\triangle(cs_i) = (\frac{f_{seq}(cs_i)}{Max(f_{seq})} + \frac{f_{act}(cs_i)}{Max(f_{act})} + \frac{f_{time}(cs_i)}{Max(f_{time})})/3$, $\overline{I}$ can be updated as following:

$$\overline{I}(cs_i.int) = \overline{I}(cs_i.int) + \alpha \cdot \triangle(cs_i)$$

where $i \in [1, |ACs|]$ and α is coefficient controlling how fast the reading behaviors could change $\overline{I}$. In *TeenRead*, we set α to 0.1. Afterwards, we should normalize $\overline{I}$ to 1 again.

B. Maintenance of Users Stress Easing Effect Vectors

$\overline{E} = (c_1.ease, c_2.ease, \cdots, c_{|SC|}.ease)$ reflects the stress easing effects of categories to the user. Each element c_i (where $1 \leq i \leq |SC|$) of $\overline{E}$ is a category in SC, and $c_i.ease \in [0, 5]$ is the easing effect of category c_i for the user, where $c_i = 0$ indicates no stress easing effect at all, and $c_i = 5$ indicates a very strong effect. Initially, $\overline{E} = (0, 0, \cdots, 0)$.

The update of $\overline{E}$ is manually conducted by users. While using *TeenRead*, users can update their stress categories and the corresponding stress level according to their current psychological states. Each time the update occurs, *TeenRead* updates user's $\overline{E}$ simultaneously. Simply, we could assign $c_i.ease$ with the same value of user's new stress level. But it stands to reason that the same c_i, in which user's stress level declines from 5 to 2, provides a better stress easing effect than that declines from 3 to 2. So the new $\overline{E}$ is expected to not only indicate user's present stress, but also reflect the changes of stress and provides a more comprehensive representation for user's stress status.

Assume the user's last stress level PL_1 = (d_1, d_2, ..., $d_{|SC|}$). After an update, stress level becomes PL_2 = (ℓ_1, ℓ_2, ..., $\ell_{|SC|}$). The change of stress level is $\triangle P = PL_1 - PL_2$ = $\{\triangle \ell_i\}$, $\triangle \ell_i$=$d_i - \ell_i$, $i \in [1, |SC|]$. To combine user's present stress and it's historical changes, as well as make sure the maximum of $c_i.ease$ doesn't exceed 5, we add this delta item to $c_i.ease$: $c_i.ease = \ell_i + \lambda \cdot \ell_i \cdot \frac{\triangle \ell_i}{5}$, where λ is an update coefficient controlling the weight of historical stress in new $\overline{E}$ and is set to 1 in *TeenRead*. In order to fuse with $\overline{I}$, $\overline{E}$ also needed to be normalized to 1 in the same way as the normalization of $\overline{I}$.

V. User Study

To test the stress easing effects and interestingness of articles recommended by *TeenRead*, we conducted an user study. In consideration that *TeenRead* is in internal test and not open to the public for now, the user study is only on 10 volunteers, but the volunteers varies from the senior undergraduates to the senior doctoral students. To meet the different needs of volunteers, we have prepared 11 reading themes, including career planning, the art of interviewing, beauty of sports and so on, from which volunteers are required to choose six most interested themes to read. The recommendation is updated every 7 days, which requires volunteers to switch themes to read every 7 days. Each time a theme is read, the user fill out the questionnaire about the stress easing effects and attractiveness of the recommended articles. After six weeks' tracking survey, we collected 45 valid questionnaires. From the study statistics, we find a significant decrease in stress level after a period of reading, which is up to 22% (1.11 compared to the maximum level 5). Moreover, the interest degrees of the four recommended articles all exceed 3.22 (the maximum degree is 5) and the stress easing effects all exceed 2.89 (the maximum degree is also 5), showing the great performance of *TeenRead* in dealing with people's psychological problems as a method a bibliotherapy.

VI. Conclusion

In this paper, we address the problem of traditional bibliotherapy in coping with stress of adolescents and propose an online reading recommendation system as a new exploration of bibliotherapy. The major advantage of *TeenRead* is that articles with good stress easing effects as well as much attraction are automatically recommended to users without human intervention. The user study has also demonstrated the good effectiveness of *TeenRead* as a new method of bibliotherapy, in which users' stress level decreases by 22% as well as users' interests in reading is pretty high.

Acknowledgment

The work is supported by National Natural Science Foundation of China (61373022, 61532015, 61521002, 71473146, 2016ZD102) and Chinese Major State Basic Research Development 973 Program (2015CB352301).

References

[1] Peilin Li and Jiangli Hao. The application of bibliotherapy for the mental health education of college students. *Beijing Education*, Z1:72–74, 2010.

[2] Boxing Zhao. A history, present condition and development of bibliotherapy. *Libray*, 3:13–15, 2003.

[3] E. Rich. User modeling via stereotypes. *Cognitive Science*, 3(4):329–354, 1979.

[4] Fen Li, Yinman Chen, and Jianxin Zhang. Review of the bibliotherapy study in china. *Journal of Modem Information*, 30(2):173–177, 2010.

[5] Peli-Hasz Elizabeth, Szilagyi Simon MD, and Baranyai Szilvia Ma. E-bibliotherapy, computer based bibliotherapy c development perspectives in relation to the effectiveness, reliability and economy. In *Proceedings of 55th ISERD International Conference*, pages 51–56, Houston, USA, December 2016.

[6] Wilawan Phornphatcharaphong. E-bibliotherapy system : Book contents for improving quality of youth's life. *Tem Journal*, 1(3):192–199, 2012.

[7] Joeran Beel, Bela Gipp, and Stefan Langer et al. Research-paper recommender systems: a literature survey. *International Journal on Digital Libraries*, 17(4):305–338, 2016.