HyKCBookRec: A Hybrid Framework for Knowledge Centric Book Recommendation Using Integrative Semantics and Variational Learning

Gerard Deepak[1], Saran B M[2], Ishdutt Trivedi[3], Saran Kaarthik B T[2] , Shiv Shankar K[2], Gokulraaj[2], Sheeba Priyadarshini J[4],Santhanavijayan A[2]

[1,2]Department of Computer Science and Engineering
[1]Manipal Institute of Technology Bengaluru, Manipal Academy of Higher Education, Manipal, India
[2]National Institute of Technology, India
[3]Trinkerr Learning Technologies, Bangalore, India
[4]CHRIST (Deemed to be University), Bangalore, India
[1]gerard.deepak.cse.nitt@gmail.com

Abstract

In today's advanced and fast running world, there is an enormous availability of digital books on the internet. But getting a proper book suggestion is still a question mark. So, several machine learning algorithms are being developed for addressing this prevailing problem. Therefore, research is being conducted to improve the recommendation model to enhance the diversity in recommendation. In this paper, a Hybridised Knowledge Centric Book Recommendation model is proposed using the hybridization of SemantoSim measure, Kullback-Liebler divergence with optimization done by novel cultural algorithm. The model has incorporated structural topic modelling, named entity recognition to uncover the related documents from the internet. The static domain ontology is aligned with upper ontology using Lin similarity measure. Indexes are then matched with the ontological terms and crawled from google books metadata repositories. The proposed system HyKCBookRec also has proven better performing model than BFRE, CFJS, CBRF, PBRDL models with an overall precision and Recall of 96.72% and 98.87% respectively.

Keywords: Book Recommender, Integrative Semantics, Knowledge Centric, SemantoSim, Variational Learning

1 INTRODUCTION

Recommender systems are used in a lot of services. The increase in development of Internet technology has increased the number of websites selling online books and competition between them is intensifying. This online book selling website helps you buy books online using recommender systems, which are one of the most powerful tools for increasing profits and retaining buyers. Both online entertainment and e-commerce companies are trying to retain their customers by allowing them to access their websites in a more personalized way. Therefore, it provides additional recommendations based on the user's past activity. This technology has proven to be extremely helpful in increasing sales. The World Wide Web is transforming into Web 3.0, an information-dense, knowledge-centric Web. Given the information density of the Semantic Web, ensuring e-learning requires a semantically justified strategy. Lack of knowledge-centric metadata-driven ontology-based model for Semantic Web 3.0 compatible and compliant book recommendations is the factor. Data 3.0 translates into a structurally cohesive and highly condensed Semantic Web. Starting from these reasons for high structural density and cohesiveness, a structural model dominated by the semantic similarity of deep learning classifications is proposed.

Contribution: A recommended model for a hybridized knowledge-centric book focusing on query-driven ontology is proposed using the Kullback-Leibler branch with cultural algorithms, SemantoSim measures, and optimizations with new cultural algorithms. The proposed HyKC BookRec system is a more powerful model than the other models and has proven to have overall Accuracy and Recall of 96.72% and 98.87%, respectively.

Organization: The remaining paper is organized as follows. Section 1 outlines the relevant Related Works. Section 2 describes the Proposed Architecture that describes structure of system. Section 4 depicts the Implementation and Section 5 describes results and performance evaluations. Finally, the paper is concluded in Section 6.

2 RELATED WORKS

Nachiapan et al., [1] address the shortcomings of existing recommended models that do not adapt well to the relatively created nature of the World Wide Web. Matthew et al., [2] provides a book suggestion model that makes use of a combination of functional algorithms to provide a better model. In [9-10], a hybrid framework with three simultaneous modules was discussed to improve the recommendation process. Anwar et al., [3] describes traditional machine learning algorithms and their categories. Ifada et al. [4] also endorses a framework that addresses the shortcomings of data sparseness commonly associated with collaborative filtering approaches and uses probabilistic keyword methods to improve recommended performance. Devika et al., [5] used a pattern mining algorithm that overcomes the shortcomings of traditional apriori. Zhang et al., [6] have initiated a well-founded nominating algorithm to improve book recommendations. Puritat et al., [7] comes up with an algorithm consisting of using multiple features to score both qualitative and quantitative data using support vector machines. Wadikar et al., [8] offers a topic-based platform for book recommendations using the Convolutional Neural Network (CNN).

Sohail et al., [11] applied a location-based sorting technique to sort the university's ranks for book recommendations. Rana et al., [12] submitted a Jaccard-like collaborative filtering (cf) to provide more accurate results. Wang et al., [13] focuses on developing a clustering-based reinforcement learning model to overcome the shortage problem and then handle noise reduction as a managerial process. Yang et al., [14] have proposed a naive Bayes algorithm for examining a user's current personalized potential demand for books. Tewari et al., [15] presents a model based on opinion mining and naive Bayes classifiers, and suggested to people the highest rated books. Xin et al., [16] have bid Linear mixing of a set of CF algorithms to increase results and perform better than the individual set filter algorithms. In [17-22] several ontological and semantic models in support of the proposed framework have been depicted.

3 PROPOSED SYSTEM ARCHITECTURE

An evolutionary algorithm known as a "cultural algorithm" is one that is motivated by sociocultural development. It consists of a belief and a population space, and a pact that sanctions communication among the origins of information. While intelligence formed in the population space is directed to the belief space, the aggregate information from many sources is combined to impact the actions made by individual agents in solving problems. Most cultural algorithms deal with numerical function optimization problems. It is predicated on the idea of the fittest surviving.

RNNs are neural networks that are a unique way to represent sequential data. A typical RNN has three states: input, hidden state, and output. The unique feature of RNNs is that they are designed to remember precise information about a sequence in their hidden state. Prior inputs are used by RNNs to impact the current input and output. The output of recurrent neural networks is determined by the sequence's previous components. Recurrent networks also have the advantage of sharing parameters across all layers of the network. By backpropagating through time, RNNs learn their weights and biases. One-to-one, one-to-many, many-to-one, and many-to-many RNNs are the four varieties. RNNs are mostly utilised in field of Natural Language Processing.

$$\frac{2 * ResnikSimilarity(c1, c2)}{IC(c1) + IC(c2)} \quad (1)$$

The Lin Similarity Measure is Node-based Measure based on the information richness of the least common subsume. It is defined in Equation (1).The SemantoSim measure, which is derived out of normalised pointwise common information measure, which is a semantic similarity metric. Equation (2) depicts the expression for SemantoSim. The Kullback–Leibler divergence is a statistical distance that tells us about the relative distance of distribution P from distribution Q. The relative entropy from P to Q for discrete probability distributions based on the same probability space is defined in Equation (3).

$$\text{SemantoSim}(x,y) = \frac{pmi(x,y) + p(x,y)log([p(x,y)])}{[p(x)*p(y)]+log([p(y,x)])} \quad (2)$$

$$D_{KL}(P||Q) = \sum P(x)\log(P(x)/Q(x)) \quad (3)$$

For distributions that are continuous random variable, relative entropy is defined in Equation (4)

$$D_{KL}(P||Q) = \int_{-\infty}^{\infty} p(x)log(p(x)/q(x)) \quad (4)$$

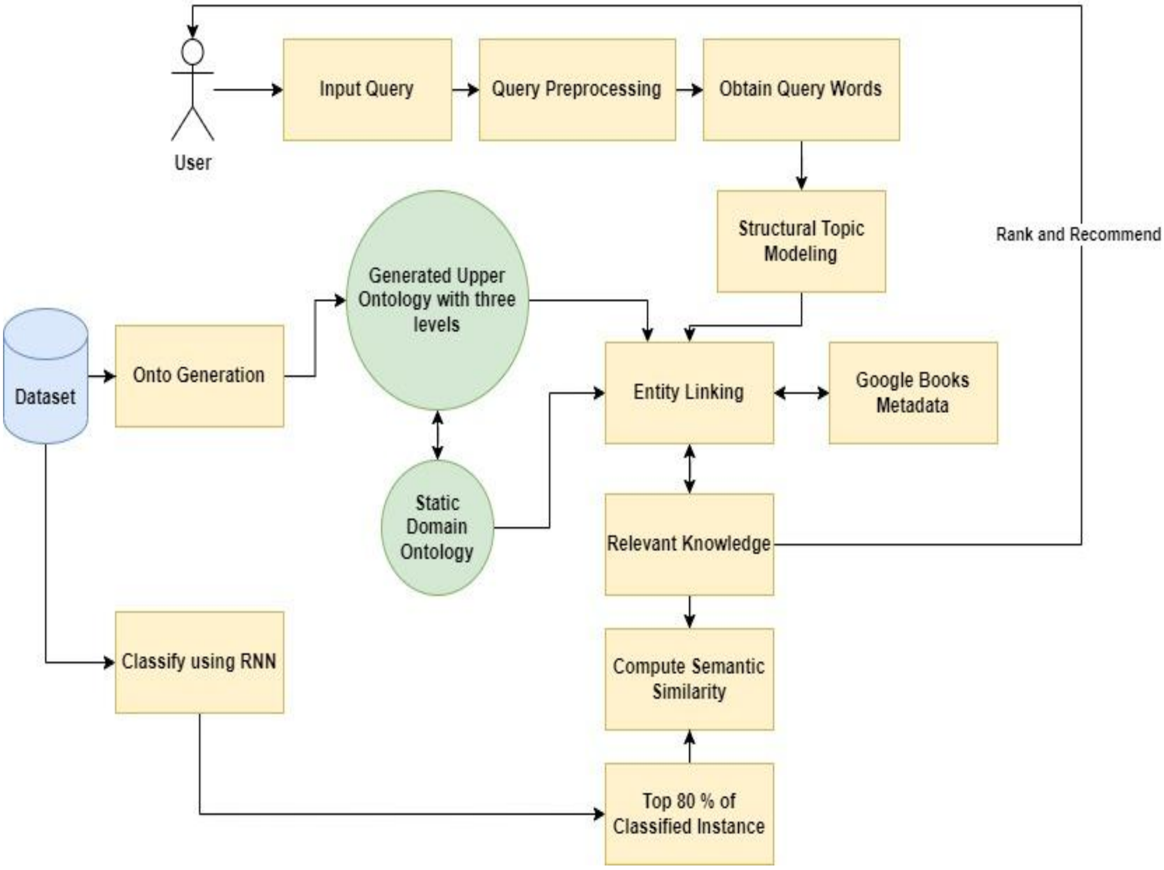

Fig.1. Proposed Architecture of HyKCBookRec Framework

The proposed HyKCBookRec is a hybridized knowledge centric semantically inclined book recommendation framework, which is driven by the user query. The user preferences are taken as input and are subjected to pre-processing. Pre-processing involves lemmatization, stop word removal, tokenization and also Named Entity Recognition (NER). At the end of the pre-processing phase, individual query words are derived, and these are subjected to Structural Topic Modelling (STM). Structural Topic Modelling is mainly achieved in order to assimilate the entities which are hidden and uncovered from the subsequent surrounding document purpose which is crawled from the vicinity of relevant topics from the world wide web structural index information. Then the dataset is subjected to ontology generation using OntoCollab and Stardog frameworks. Once the ontology is generated, it is ensured that the ontology is only an upper ontology of only three levels. Three-level ontology is incorporated because detailed ontology can deviate the relevance from the dataset. However, static domain ontologies consisting of 1284 instances are also modelled based on manual modelling using web prodigy and customized crawler-based entities. The static domain ontology is first aligned with the generated upper ontology using lin similarity measure. The aligned ontologies are further sent for entity linking. Entity linking happens with ontology alignment but using the SemantoSim measure with a threshold of exactly 0.5 because we require a greater number of entities to pass through this phase. Entities are linked between the matched ontological terms and the indexes crawled from Google Books Metadata repository. After this phase, the relevant set is formulated which is a space-oriented structure which houses all the relevant entities. The dataset is now classified using Recurrent Neural Networks (RNN). RNN being a deep-learning classifier it automatically classifies the dataset using the auto handcrafted feature selection methodology. Among the automatically

274

obtained classes, top 80% of the classified instances under each class and category is selected

The principal class and the individual class are randomly selected and are subjected to semantic similarity computation with the relevant knowledge set. To do this haphazardly, an agent is written using AgentSpeak. In this case, for the SemantoSim measure we take the threshold as 0.75 and for the KL divergence we take 0.25 as the step deviation. The reason for using both KL divergence and SemantoSim measure is to ensure that the strength of relevance computation is maintained because this participates in the final recommendation. The cultural search algorithm acts as an active meta heuristic module, where the feasible relevant sets are transformed into the most relevant optimal solution set by keeping KL Divergence and SemantoSim measure as objective functions. The entities which are relevant are used for loading the e-books from the dataset repository and yielded subsequently to the user. If the user is satisfied the search halts here, else the process continues until the user is satisfied.

4 IMPLEMENTATION

The proposed book recommendation framework was implemented in Python 3.10.5 using Google's Collaboratory as the development engine on an Intel core I5 processor with 32GB DDR4 2666MHz RAM clock speed and an external plugin supported GPU. The preprocessing task was done using python's NLTK library. The experimentations are conducted on an integrative customized dataset of 4 standard datasets namely the Chalchitra Talks Book recommendation dataset, GetData.IO-Book Recommendation for Entrepreneurs & Investors dataset, Goodbooks-10k dataset, Users-Books-Dataset of Real World were used as individual datasets but were integrated at a common point. If integration was not possible, they were annotated and integrated at one common point. Wherever there was a possibility of integration if there were no common points they were just left as it is, they were annotated and converted into .csv and made into integrative single large data set for experimentation. The proposed HyKCBookRec Algorithm is depicted as Algorithm 1.

Algorithm 1: Proposed HyKCBookRec Algorithm

Input: User query(q), static data ontology(d_o) relevant to the dataset,
metadata crawled from google books(gm_d), Dataset(d_s)
Output: E-books which are relevant to the query

begin
Step 1: Preprocessing is performed on the input, user query(Q,w), which involves tokenization, lemmatization, stop word removal, and NER.

Step 2: for each query(qw), apply STM
(i.e) Set S t <- Subject it to STM

Step 3: Parsed dataset categorizes and generates Ontology using OntoCollab and Star Spacedog framework

Step 4: for each onto O.next() != NULL:
Load static domain ontology SDO
Align (SD0 with O.ele()) until O.ele()!=NULL using Lin Similarity

Step 5: Entity linking is based on SemantoSim computation between the entities that are aligned between static domain ontology and entities obtained from the google books metadata

Step 6.1: Initialize the starting population space
6.2: Initialize optimal space
6.3: Loop until the termination condition is satisfied
6.4: Individual tasks in population space performed
6.5: Using the K-L divergence step ratio and the SemantoSim measure, use semantic similarity under cultural search algorithm as fitness functions
6.6: Evaluate each individual using the fitness functions
While (SemantoSim measure<7.5 and K-L divergence step ratio < .25)
6.7: Produce a new generation of children by choosing the parents
6.8: Using the influence function, the structure of the children is changed by ideal space
6.9: With the accept function, the best individuals affect the optimal space which in turn updates the ideal space
6.10: With this use SemantoSim ratio and K-L divergence objective functions under multiple agents called user AgentSpeak

Step 7: Yield the matching instances and send them to the user.

5 RESULTS AND PERFORMANCE EVALUATION

The effectiveness of the suggested HyKCBookRec which is a Hybridised knowledge instilled approach for book recommendation was computed using F-measure, Precision, Accuracy, Recall, and False Discovery Rate (FDR) as standard and preferred metrics. These metrics demonstrate the applicability of the HyKCBookRec framework's outcomes. FDR shows or quantifies the number of false positives yielded by the framework. More the value of accuracy, f-measure, recall, and precision better the framework's performance. Lesser the value of FDR better the performance of the framework.

NeuroQuantology | OCTOBER 2022 | VOLUME 20 | ISSUE 12 | PAGE 272-278| DOI: 10.14704/NQ.2022.20.12.NQ77024

Table 1 .Evaluation of the proposed HyKCBookRec Framework

Search Technique	Average Precision %	Average Recall %	Accuracy %	F-Measure %	FDR
BRFA	87.28	90.02	88.65	88.62	0.13
CFJS	90.22	93.04	91.63	91.60	0.10
CBRF	91.12	93.72	92.42	92.40	0.09
PBRDL	92.11	94.36	93.235	93.22	0.08
Proposed HyKCBookRec	**96.72**	**98.87**	**97.79**	**97.78**	**0.04**

From Table 1, it is observed that the proposed framework performs better in quantifying and comparing HyKCBookRec – for 2247 queries for which empirical data has been gathered. Other models lag because the BFRE model uses the fussy-based aggregation Ordered Ranked Weighted Aggregation (ORWA) operator. This ordered weighted strategy does not perform well as expected. CFJS algorithm also does not perform well because it is based on Collaborative Filtering with Jaccard Similarity. Collaborative Filtering requires reading computation metrics where every entity cannot be rated perfectly. Though Jaccard Similarity is a robust similarity measure, the rating-based scheme fails to perform well. CBRF is the Clustering-Based Reinforcement Learning method. This method makes the learning and computational load high. PBRDL uses a deep learning model. It results in overfitting by incorporating a deep learning model with a very shallow auxiliary knowledge fit into the framework. With shallow auxiliary knowledge and deep learning, it leads to overfitting.

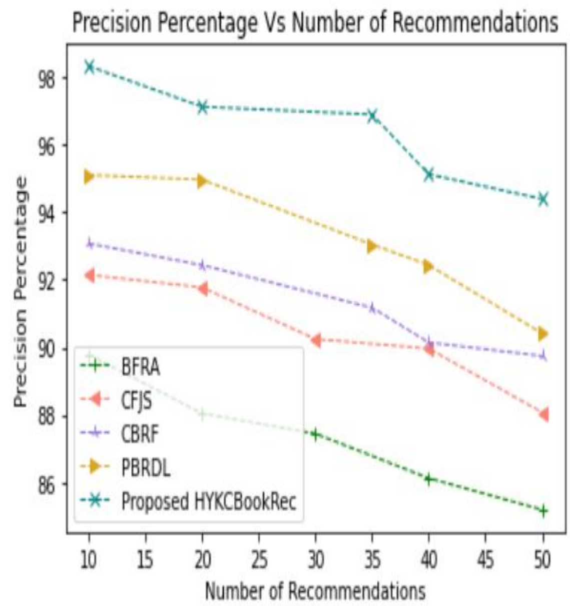

Fig.2. Precision Vs. No. of Recommendations Distribution Curve

The Precision vs. No. of Recommendations distribution curve is depicted in the above figure. The proposed HyKCBookRec occupies the first position, PBRDL occupies the second position, CBRF occupies the third position, CFJS occupies the fourth position, and the last position is occupied by BFRA. The proposed model is faster because it is semantically inclined, and several different semantic similarity measures have been used with differential thresholds. The relevance computation mechanism is quite effective.The proposed HyKCBookRec computes the optimal solution set using the novel cultural algorithm under the SemantoSim and KL divergence (Kullback–Leibler divergence) with differential thresholds. Optimal solution set derivation takes place from a feasible solution set using cultural algorithm, so this model performs better than the other models.

6 CONCLUSIONS

Recommender systems are a very powerful tool used to facilitate the user selection process. Here, based on a new cultural algorithm, we introduced a recommendation system. Cultural algorithms are used to suggest recommended models for hybrid knowledge-centric books. Kullback-Leibler divergence and SemantoSim measure with optimization is done by novel cultural algorithm. On comparing various results of the method / model we arrived at an optimal solution. The proposed HyKCBookRec system achieves 96.72% and 97.795% overall accuracy which is better than other baseline models. The main goal was to enhance the diversity and the relevance of results in recommendations without the need to register for a long time and without having a large amount of profile information or browsing history.

REFERENCES

Nachiappan, R., & Deepak, G. (2022). OSIBR: Ontology Focused Semantic Intelligence Approach for Book Recommendation. In International Conference on Digital Technologies and Applications (pp. 397-406). Springer, Cham.

Mathew, P., Kuriakose, B., & Hegde, V. (2016, March). Book Recommendation System through content based and collaborative filtering method. In 2016 International conference on data mining and advanced computing (SAPIENCE) (pp. 47-52). IEEE.

Anwar, K., Siddiqui, J., & Saquib Sohail, S. (2019, February). Machine learning techniques for book recommendation: an overview. In Proceedings of International Conference on Sustainable Computing in Science, Technology and Management (SUSCOM), Amity University Rajasthan, Jaipur-India.

Ifada, N., Syachrudin, I., Sophan, M. K., & Wahyuni, S. (2019). Enhancing the performance of library book recommendation system by employing the probabilistic-keyword model on a collaborative filtering approach. Procedia Computer Science, 157, 345-352.

Devika, P., Jisha, R. C., & Sajeev, G. P. (2016, December). A novel approach for book recommendation systems. In 2016 IEEE international conference on computational intelligence and computing research (ICCIC) (pp. 1-6). IEEE.

Zhang, F. (2016). A personalized time-sequence-based book recommendation algorithm for digital libraries. IEEE access, 4, 2714-2720.

Puritat, K., Julrode, P., Ariya, P., Sangamuang, S., & Intawong, K. (2021). Book Recommendation for Library Automation Use in School Libraries by Multi Features of Support Vector Machine. International Journal of Advanced Computer Science and Applications, 12(4).

Wadikar, D., Kumari, N., Bhat, R., & Shirodkar, V. (2020). Book recommendation platform using deep learning. International Research Journal of Engineering and Technology IRJET:, 6764-6770.

Tian, Y., Zheng, B., Wang, Y., Zhang, Y., & Wu, Q. (2019). College library personalized recommendation system based on hybrid recommendation algorithm. Procedia CIRP, 83, 490-494.

Sariki, T. P., & Guntur, B. K. (2022). An Aggrandized Framework For Enriching Book Recommendation System. Malaysian Journal of Computer Science, 35(2), 111-127.

Sohail, S. S., Siddiqui, J., & Ali, R. (2017). A novel approach for book recommendation using fuzzy based aggregation. Indian Journal of Science and technology, 8(1).

Rana, A., & Deeba, K. (2019, November). Online book recommendation system using collaborative filtering (with Jaccard similarity). In Journal of Physics: Conference Series (Vol. 1362, No. 1, p. 012130). IOP Publishing.

Wang, X., Wang, Y., Guo, L., Xu, L., Gao, B., Liu, F., & Li, W. (2021). Exploring Clustering-Based Reinforcement Learning for Personalized Book Recommendation in Digital Library. Information, 12(5), 198.

Yang, W. (2022). Personalized Intelligent Recommendation Algorithm Design for Book Services Based on Deep Learning. Wireless Communications and Mobile Computing, 2022.

Tewari, A. S., Ansari, T. S., & Barman, A. G. (2014, November). Opinion based book recommendation using naive bayes classifier. In 2014 International Conference on Contemporary Computing and Informatics (IC3I) (pp. 139-144). IEEE.

Xin, L., Haihong, E., Junde, S., Meina, S., & Junjie, T. (2013, December). Collaborative book recommendation based on readers' borrowing records. In 2013 International Conference on Advanced Cloud and Big Data (pp. 159-163). IEEE.

Arulmozhivarman, M., & Deepak, G. (2021, March). OWLW: ontology focused user centric architecture for web service recommendation based on LSTM and whale optimization. In European, Asian, Middle Eastern, North African Conference on Management & Information Systems (pp. 334-344). Springer.

Krishnan, N., & Deepak, G. (2021, May). KnowSum: knowledge inclusive approach for text summarization using semantic alignment. In 2021 7th International Conference on Web Research (ICWR) (pp. 227-231). IEEE.

Adithya, V., & Deepak, G. (2021, March). OntoReq: an ontology focused collective knowledge approach for requirement traceability modelling. In European, Asian, Middle Eastern, North African Conference on Management & Information Systems (pp. 358-370). Springer.

Varghese, L., Deepak, G., & Santhanavijayan, A. (2021, January). A fuzzy ontology driven integrated IoT approach for home automation. In International Conference on Digital Technologies and Applications (pp. 271-277). Springer, Cham.

Pushpa, C. N., Deepak, G., Thriveni, J., & Venugopal, K. R. (2016). A hybridized framework for ontology modeling incorporating latent semantic analysis and content based filtering. International Journal of Computer Applications, 150(11).

Srivastava, R. A., & Deepak, G. (2021). PIREN: prediction of intermediary readers' emotion from news-articles. In Data Science and Security (pp. 122-130). Springer, Singapore.