RESEARCH ARTICLE OPEN ACCESS

"Review On Document Recommendations Using HHN and RWR Algorithms"

Mr. Umesh A. Patil (Assistant Professor), Miss Vaishnavi C. Hiwarekar (Student), Miss Sayali H. Gholase (Student), Mr. Ketan T. Khot (Student), Miss *Poonam S. Patil (Student)*
(Department of Computer Science and Engineering, D. Y. Patil Technical Campus Talsande, Kolhapur)

ABSTRACT
The advancement of several significant technologies, such as artificial intelligence, cyber intelligence, and machine learning, has made big data penetrate the industry and academic field and our daily life and a variety of cyber-enabled applications. This article focuses on a deep correlation mining method in heterogeneous big data environments. A hierarchical hybrid network (HHN) model is constructed to describe multitype relationships among different entities.

A series of measures are defined to quantify the internal correlations within one specific layer or external correlations between different layers. An intelligent router based on a deep reinforcement learning framework is designed to generate optimal actions to route across the HHN. An improved random walk with the restart-based algorithm is then developed with the intelligent router, based on the hierarchical influence across networks associated with multiple correlations.

An intelligent recommendation mechanism is finally designed and applied to support users' collaboration works in scholarly big data environments. Experiments based on DBLP and ResearchGate data show the practicability and usefulness of our model and method.

Keywords - Correlation mining, cyber intelligence, heterogeneous big data, hierarchical hybrid network (HHN), reinforcement learning, social influence

Date of Submission: 09-04-2022 Date of Acceptance: 26-04-2022

I. INTRODUCTION

The development of emerging technologies has enabled big data to quickly penetrate both industry areas.

Some recommendation systems consider multiple relationships, they measure these relationships separately, rather than combining them in an associative way. When dealing with different types of entities in a constructed network model, most studies calculate the relevance among the same type of entities in a non-discriminatory way, and thus, do not take multilevel factors into account.

II. LITERATURE REVIEW:

In the cyber social computing environment, network modelling and correlation or relationship analysis among users have become fundamental and even indispensable elements of application development.

A two-stage discriminant model was constructed to model the group-oriented decision-making process.

Utilizing a multi-stage model based on users' preferences and latent social connections, the dynamic influence of mutual influence could be predicted to be more accurate. We used three factors: mobility influence, content similarity, and social relationship together to identify the on-site user in social networks.

Now we are using the hierarchical hybrid network and Random walk with restart techniques to improve the system performance very accurately and predict the recommended result.

III. OBJECTIVES:

- To build a system to provide intelligent recommendations in the context of heterogeneous big data integration from multiple data sources.
- To build a system to provide a hierarchical hybrid network (HHN) model to describe multiple associations among different entities.
- To build a system to train the model using DBLP Datasets.
- To evaluate the system using precision and recall

IV. PROPOSED WORK:

I. HIERARCHICAL HYBRID NETWORK MODEL:

Three basic relations among users and items are considered to build the HHN model. Definitions can be expressed as follows: -

GHHN (N, E, C)

Where N = UN ∪ IN indicates a combination of users and items in the heterogeneous network model, which is basically used to form a user network UN and an item network IN. In HHN, the user network is a multi-layer structure that describes the user correlations in a hierarchical manner, and the item network describes what items are relevant or similar and what they represent.

UN = {*u*1, *u*2, . . . , *um*} represents a non-empty set of nodes (i.e., the users) in the HHN model. In particular, *ui* represents a specific user, and *ui = (*UID*i* , Int*i* , LYR*i)*, in which UID*i* is the user ID, Int*i* is a vector with a set of keywords that indicates the interests of *ui*, and LYR*i* is the specific layer that *ui* belongs to in UN.

IN = {itm1, itm2, . . . , itmn} represents a non-empty set of nodes (i.e., the items) in the HHN model. In particular, itmi indicates a specific item, and itmi = (IIDi , Fi , Cuii , Ctgi), in

Which IIDi is the item ID, Fi is a vector with a set of keywords to represent and describe the semantic features for itmi, Cuii is a vector with a set of users who are interested in itmi, and Ctgi indicates the specific category that itmi belongs to in IN.

E = (EUN, EIN, EUI) indicates a combination of edges that connect different nodes in UN and IN. In particular, it includes three important relations as: 1) EUN indicates the direct or indirect relationships between ui and u j; 2) EIN indicates one kind of semantic-aware relationship between itmi and itm j ; and 3) EUI indicates one kind of interest-based relationships between ui and item j. *C* = {*Ci j* |if *_eij _ E*} indicates a series of measures, which are used to represent the multiple correlations between different users and items within the HHN model.

II. RANDOM WALK WITH RESTART MODEL: -

An RWR model is used to extract the structure-aware features, and determine the optimal node based on their significance in a constructed HHN model.

The basic framework of the RWR model for recommendations from the HHN can be expressed as follows:

$$\mathrm{HR}(t+1) = \lambda M\, \mathrm{HR}(t) + (1 - \lambda)\, q$$

where _, ranging from 0 to 1, is a damping coefficient. HR t indicates a ranking score vector at iteration step *t*. *q* is the initial vector when starting the RWR model, which is constructed according to the initial state of the network.

Specifically, *q* is initialized as [0, 0, . . , 1, . . . , 0, 0] and we set HR0 = *q*, in which “1” indicates the target vertex *vt* at the beginning.

M is a transfer matrix, which indicates the probability of each vertex to transfer to one another.

V. ARCHITECTURE:

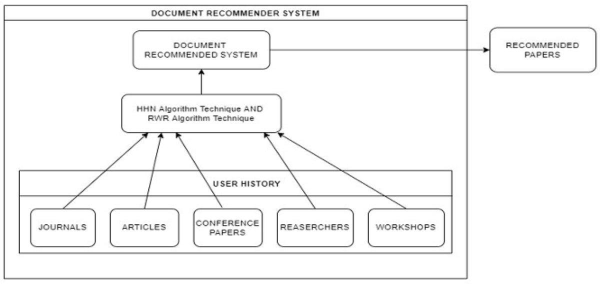

Fig: Architecture Diagram

VI. IMPLEMENTATION STEPS:

DATASET:

We are using the DBLP and Research-Gate data set. In that, we are using 70% data for training purposes and 30% data for Testing purposes.

MODULES:

1. Data Collection and Model Graph Representation
2. Built-up Hierarchical User Relation
3. Item Relation with Semantic Attribute
4. Recommendation with Random Walk with Restart

VII. PROJECT SCOPE:

Furthermore, to evaluate the effectiveness of the proposed method, especially when handling intelligent recommendations with heterogeneous data, we tested both the user-based and item-based intelligent recommendations in the experiment with scholarly big data.

Given a randomly selected target object (researcher or article) with a maximum of 50 alternatives, the proposed method was used to generate the optimal academic collaborations alternatives, and the results were compared with the aforementioned four baseline methods.

REFERENCES:

[1]. X. Wang, L. T. Yang, H. Liu, and M. J. Deen, “A big data-as-a-service framework: State-of-the-art and perspectives,” IEEE Trans. Big Data, vol. 4, no. 3, pp. 325–340, Sep. 2018

[2]. F. Xia, Z. Chen, W. Wang, J. Li, and L. T. Yang, “MVC Walker: Random walk-based most valuable collaborators recommendation exploiting academic factors,” IEEE Trans. Emerg. Topics Comput., vol. 2, no. 3, pp. 364–375, Sep. 2014

[3]. F. Ali et al., “Merged ontology and SVM-based information extraction and recommendation system for social robots,” IEEE Access, vol. 5, pp. 12364–12379, 2017.

Mr. Umesh A. Patil, et. al. “Review On Document Recommendations Using HHN and RWR Algorithms.” *International Journal of Engineering Research and Applications (IJERA)*, vol.12 (04), 2022, pp 20-22.