International Journal of
INTELLIGENT SYSTEMS AND APPLICATIONS IN ENGINEERING
ISSN:2147-6799 www.ijisae.org Original Research Paper

Analysis of Recommender Systems in Heterogeneous Information Networks using HINPy

Sadhana Kodali [1], Madhavi Dabbiru[2], Thirupathi Rao Komati[3]

Submitted: 05/12/2023 **Revised**: 14/01/2024 **Accepted**: 28/01/2024

Abstract: Recommender systems play a pivotal role in enhancing user experiences across various online platforms, from e-commerce websites to social media and content streaming services. Traditional recommender systems have primarily relied on homogeneous data structures, limiting their ability to effectively capture complex user-item interactions. Heterogeneous Information Networks (HINs) have emerged as a powerful paradigm to address these limitations by modeling diverse types of entities and relationships present in real-world recommendation scenarios. This paper provides a comprehensive review on the usage of HINPy which is a python workbench used for the analysis of recommender systems. HINPy is also a powerful workbench for the representation of networks. This paper analyses the cuisine based recommender system using HINPy and focuses on introducing the foundational concepts of HINPy and unique advantages in capturing rich and diverse information about users, items

Keywords: *Heterogeneous Information Networks, Python workbench, HINPy, recommender systems.*

1. Introduction

In today's digitally connected world, recommender systems have become indispensable tools for guiding users through the vast sea of information and content available online. These systems, ranging from e-commerce platforms to social media networks and content streaming services, play a pivotal role in enhancing user experiences by suggesting relevant items or connections based on their preferences and behaviors. Traditionally, recommender systems have relied on homogeneous data representations, which often fall short in capturing the intricate relationships and diverse information present in real-world recommendation scenarios.

Heterogeneous Information Networks (HINs)[1] have emerged as a cutting-edge approach to address the limitations of conventional recommender systems[2].

Unlike traditional systems that typically represent data as a uniform graph, HINs embrace the complexity of modern data sources by accommodating multiple types of entities (eg., users, items, attributes) and relationships (eg., user-item interactions, social connections, content associations). This diversity of data types and connections in HINs offers a more holistic view of the underlying information landscape and, consequently, opens up new avenues for making personalized and context-aware recommendations.

This paper aims to provide an in-depth exploration of the intersection between recommender systems and HINs. It delves into the foundational concepts, methodologies, and practical applications that define this evolving field. By leveraging the heterogeneous nature of data, HIN-based recommender systems promise to revolutionize recommendation technology, offering improved accuracy, serendipity, and user satisfaction.

In the following sections, an overview of HINs and their unique characteristics are discussed. This paper highlights the fusion of different data sources and the challenges associated with such heterogeneous data integration. To overcome these challenges the use of python workbench for the analysis of the heterogeneous infused data is discussed.

The paper is organized in the following manner: The latest recommender system analysis of various heterogeneous information networks are discussed in Section 2,Section 3 gives a step-by-step methodology for the installation of python workbench HINPy and using it, Section 4 gives the experimental results and the importance of HINPy is concluded in Section 5.

2. Literature Survey

2.1 Homogeneous Information Network

In the real world objects are interconnected either physically in abstraction which forms a data network called the Heterogeneous Information Network. The objects interconnected and the links between them are of different type it is called a Heterogeneous Information Network. Multiple Homogeneous Information Networks which have object inter-connectivity of the same type and a single

[1]*Assistant Professor, Department of CSE(H), Koneru Lakshmaiah Education Foundation, Vaddeswaram, Guntur Dt, Andhra Pradesh, India.* *ksadhana@kluniversity.in, Orcid Id: 0000-0001-8744-6184.*
[2]*Professor,Dr. Lankapalli Bullayya College of Engineering ,Visakhapatnam,Andhra Pradesh,India.*
[3]*Professor, Department of CSE, Gitam deemed to be University, Visakhapatnam, Andhra Pradesh, India.*
* *Corresponding Author Email: ksadhana@kluniversity.in*

relationship that exits between these objects are projected in the Heterogeneous Information Networks. Therefore a Heterogeneous Information Network is a collection of multiple Homogeneous Networks.

When the information is fused to form the Heterogeneous Information Networks a lot of valuable knowledge can be extracted from these heterogeneous Networks. Mining of the fused knowledge from the Heterogeneous Information Networks is a challenging task using traditional algorithms[3]. The reason for complexity is to extract the semantically lying links which can lead to hidden object relationships which may be essential for proposing the new patterns.. Identifying the new patterns may have a remarkable change for proposing new recommendations and can change the trend of user buying strategies or improve the problem solving capability in complex networks by paving a way to identify the solution for an existing problem in an optimized approach.

In a Heterogeneous information network, the nodes (V) can represent different types of entities, and the edges (E) can represent various types of relations between these entities. More clearly V represents entities (nodes), E represents relations (links), and there are mapping functions f_v and f_e. The mapping functions[4] f_v and f_e play a crucial role in associating entities and relations with their respective types, providing a structured representation of the diverse elements in the network.

The introduction of heterogeneous information networks enriches the modeling capacity of the graph, allowing for a more nuanced and comprehensive representation of complex relationships within a system. This framework finds applications in various domains, such as recommendation systems, social network analysis, and knowledge graphs, where entities and relations can be inherently diverse and multi-faceted.

By considering multiple entity types or relation types, heterogeneous information networks offer a more realistic and flexible modeling approach, capturing the intricacies of real-world systems where entities and relations are not uniform.

The recommender system can be conceptualized as a link prediction task within the framework of an information network. While some traditional recommender systems rely on bipartite graph modeling, these approaches face challenges in effectively incorporating various auxiliary information. Although some methods have been proposed for integrating specific types of auxiliary information, they often lack universality.

In recent years, HIN-based recommender systems have emerged as a successful solution to the challenges posed by modeling heterogeneous auxiliary information and user interaction behaviors in a unified manner. This approach not only addresses issues related to data sparsity and cold start problems in recommender systems but also enhances the interpretability of these systems. Consequently, HIN-based recommender systems have garnered widespread attention and found applications across diverse domains.

The HIN-based recommendation process involves two key steps: Firstly, the construction of a heterogeneous information network based on user-item interaction data and all available auxiliary information. Subsequently, a recommendation model is designed specifically to accommodate the characteristics of the heterogeneous information network. This two-step approach contributes to the adaptability and effectiveness of HIN-based recommender systems, offering a more holistic and interpretable solution to the challenges faced by traditional recommendation approaches. By unifying diverse information sources within a comprehensive network structure, HIN-based recommender systems provide a powerful means to enhance recommendation accuracy, particularly in scenarios characterized by data sparsity and cold start issues. Therefore these systems have become an integral part of the recommender system landscape, demonstrating their versatility and effectiveness in real-world applications. Heterogeneous Information Networks (HINs) can be constructed not only from structured data but also from unstructured data that has been processed through techniques like entity extraction and relation extraction. By incorporating both structured and unstructured data into the construction of HINs, recommendation systems can benefit from a richer representation of information, leading to more contextually aware and accurate recommendations.

Recommender systems in Heterogeneous Information systems can be classified based on taxonomy. The taxonomy based models are further classified as: similarity measurement, matrix factorization, Graph representation learning predefined meta-paths. The similarity between two entities is measured by the probability of reaching similar entities during the random walk proposed in [6] as PCRW.

2.2 Taxonomy based Models:

Fig 1: Taxonomy based Recommender Systems.

Taxonomy based models can be classified based on the Figure 1.

Similarity Measurement:

a. Relation Based Models: The Relation based model makes use of random-walks and RHSN [7] is an example of this model. In which the authors considered the recommendations as a ranking model using the random walks.

b. Meta-path based Recommender Systems: There are three common types of meta-path-based similarity metrics shown in Figure 1.

1.Path Count: This involves counting the number of common meta-paths between two entities. Entities that share a higher number of common meta-paths are considered more similar. The best example is Path Sim[5].

2.Path-based Random Walk: This metric involves simulating random walks on the graph.

3.Path-based Pairwise Random Walk: This is an extension of path-based random walk where the focus is on pairwise comparisons between entities. It considers the probabilities of reaching common neighbors during the random walk for each pair of entities.HeteSim[7] uses a path based pair wise random walks.

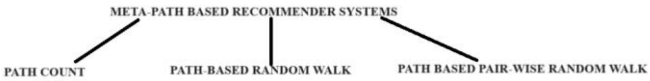

Fig 2: Recommender Systems in HIN based on meta-paths.

c. Matrix Factorization based recommender system:

The authors in [8] proposed a matrix factorization model for items which lack connectivity. The hidden vectors are identified and the recommendations are done based on the hidden vectors. The variants of the matrix factorization model are regularization based which use similarity measures in the matrix to determine the hidden vectors.HeteSim [9] is a well-known regularization based method. Another approach for Matrix Factorization is Neural Matrix Factorization which uses the concepts of neural networks and deep learning.

d. Graph Representation Learning: The traditional Neural Networks achieved the best recommendation results but they failed to handle the graph based structures.

Two stage training based approach use per-training and fine tuning methods for embedding the tasks of recommendations. MoHINRec [11] is an example of this model. The drawback of two stage training based recommender system does not use the supervised information where as the end-end training approach uses a supervised information.

3. Methodology

With the innovation of the python work bench for the analysis of the recommender systems using Heterogeneous Information Networks, the analysis has become more effective. HINPy is a python work bench which can study the fusion of Information and analysis of recommender in Heterogeneous Information Networks. HINPy is also used for network representations and domain representations like ecology, social network analysis, medical networks etc.

HINPy can represent topological data where objects are ontologically connected by different relationships or different entities which are connected. HINPY workbench has many uses like extracting the metrical structures which means similarities and distances between the objects,embedding of spaces in relative order known as hilbertian structures and analyzing empirical analysis ,distributions of Bayesian structures. To use this workbench first HINPy needs to be installed. This can be done by using the command for installation.

Once HINPy workbench is installed import HINPy to the work space. The Information Network which needs analysis can be prepared as a csv file and the relationships between the entities can be named in the csv file. The relationships and the entities can be named for the recommender system according to the context. Figure 1 indicates the installation of the python workbench HINPy in google colab.

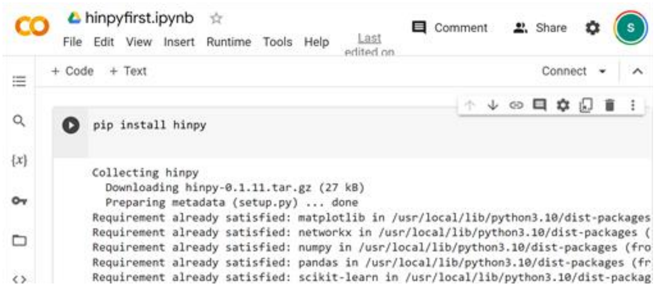

Fig 3: Installing HINPy using google colab.

3.1 Preperation of the Dataset:

The Dataset for the analysis of recommender systems using HINPy needs to be prepared with the relationship between the objects. The entities and relationships for the user cuisine recommender system is considered for the experimentation and the dataset is prepared. The relation_name followed by the group of entity,name of the entity,group of entity 2,name of entity 2.

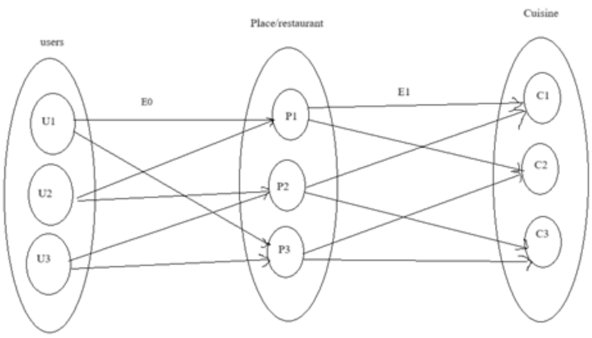

Fig 4: Object groups and objects identified.

The Figure 4 is used for the analysis of user-cuisine recommender system where the first group indicates the user groups a and all the user entities fall in this group. The second group indicates the places or restaurants visited by the users. The third group indicates the cuisines offered in the corresponding restaurants. There exists an edge E0 from the users group which is a directed edge from the user to the restaurant that he visited. The figure also shows a directed edge E1 from the place/restaurant to the cuisine.

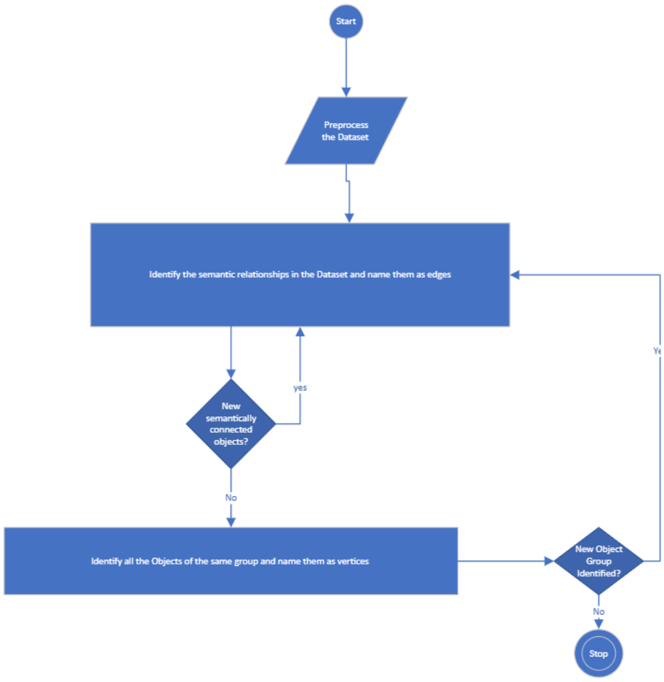

Fig 5: Data Preparation for using HINPy

The Figure 5 indicates the steps for preparing the dataset before uploading the file and using it.

4. Experimental Results

To make use of the python workbench HINPy it can be first installed. For the conduction of the experiments Google Colab note book is used. The HINPy should be imported to the workspace using the statement:

import hinpy

Load the dataset using the statement

hin=hinpy.HIN(filename="/content/Dataset.csv")

Description of the Dataset:

The Dataset Restaurant and Consumer data [12] is obtained from UCI Machine Learning repository and the authors used this dataset in [13].The dataset is used for recommendations purpose which contain the user ratings of the place visited, cuisine ordered and the service rating. The data set consists of 1161 instances with user-id and place-d as nominal instances and various ratings of cuisine,place and service are numeric instances. The data set needs to be per-processed before giving it as an input by identifying the names of the semantic relationships and by maintaining equal number of instances for all the possibilities. After pre-processing and identifying all the possible object groups and their relationships the total number of instance are 2232.The network schema of this data set is identified as the following: has_opted, users, userID, place, placeID .Where has_opted is the relationship name from the group of users 'u ',with corresponding user_ids ,from a group of places 'p' with corresponding place_ids.

We can check the corresponding probabilities of the entities using the method GetObjectGroupPositionDic() shown in Figure 6.

```
hin.GetObjectGroupPositionDic('u')
{'U1119': 0,
 'U1041': 1,
 'U1014': 2,
 'U1048': 3,
 'U1059': 4,
 'U1065': 5,
 'U1033': 6,
 'U1044': 7,
 'U1030': 8,
 'U1038': 9,
 'U1020': 10,
 'U1108': 11,
 'U1114': 12,
 'U1034': 13,
 'U1015': 14,
 'U1069': 15,
 'U1009': 16,
 'U1128': 17,
 'U1010': 18,
 'U1086': 19,
 'U1031': 20,
 'U1012': 21,
 'U1036': 22,
```

Fig 6: Group Object probabilities.

All the link groups can be recognized using the method GetLinkGroupsNames().The Figure 7 indicates the possible link groups identified between users who visited the restaurants, recipes ordered by the users at various restaurants.

```
hin.GetLinkGroupsNames()
['has_visited', 'E0', 'E1', 'inverse_has_visited', 'inverse_E0', 'inverse_E1']
```

Fig 7: Group Links identified in the dataset.

The Proportional abundance in the network schema can be measured using the method proportional_abundance().

The proportional abundance measured between the objects are also called as the apportionment which are obtained by the random walks on the meta-paths and are the probability mass functions on the random walks of the various meta-paths present. Proportional Abundance is used for understanding the proportion or percentage of each node type in the overall network. It gives insights into the relative importance or prevalence of different entity types. The proportional_abundances() with respect to 'E0' and 'E1' are indicated in Figure 8.

```
[11] ar=hin.proportional_abundances(path=['E0','E1'])

print(ar)

    (0, 5)       0.14285714285714285
    (0, 10)      0.8571428571428571
    (1, 35)      0.16666666666666666
    (1, 5)       0.16666666666666666
    (1, 46)      0.10416666666666667
    (1, 13)      0.0625
    (1, 3)       0.125
    (1, 10)      0.25
    (1, 49)      0.125
    (2, 35)      0.125
    (2, 10)      0.25
    (2, 32)      0.0625
    (2, 18)      0.0625
    (2, 46)      0.25
    (2, 27)      0.125
    (2, 42)      0.125
    (3, 46)      0.03333333333333333
    (3, 35)      0.03333333333333333
    (3, 5)       0.13333333333333333
    (3, 49)      0.19999999999999998
    (3, 10)      0.5
    (3, 23)      0.09999999999999999
    (4, 35)      0.18749999999999997
    (4, 5)       0.0625
    (4, 10)      0.25
```

Fig 8: Proportional Abundance of E0 and E1.

Diversity often refers to the variety or heterogeneity of nodes and edges within the network. It could encompass different types of entities, relationships, or attributes associated with them.

Mean Diversity is a measure that averages the diversity across the various types of nodes or edges in the network. It provides a summary statistic indicating how diverse the information is on average.

Calculating mean diversity in a heterogeneous information network involves assessing the distribution of different types of nodes or edges and then computing a measure of the overall diversity. The measure of diversity where various users visit various restaurants and order various cuisines. This can be useful for understanding the overall complexity and richness of information in the network.

Analyzing collective diversity in HINs is valuable for understanding the overall complexity of the network and identifying patterns that might not be apparent when focusing only on individual node or edge types. It can provide insights into the holistic information landscape within the network, aiding in tasks such as recommendation systems, community detection, and information retrieval.

Computing the Collective Diversity between E0 and its inverse relations which is shown in Figure 9.In the case of arbitrarily complex meta-paths the collective_diversity() method can be used with respect to inverse relations. The mean_diversity of E0 and E1 shows 4.4252 where as the collective_diversity calculated for E0 and E1 is 8.2049.

```
[15] hin.collective_diversity(['E0','inverse_E0'])
     133.70735005558487

[16] hin.collective_diversity(['E1','inverse_E1'])
     729.6724982987545

[27] hin.mean_diversity(['E0','E1'],alpha=2.0)
     4.425200668913451

[29] hin.collective_diversity(['E0','E1'],alpha=2.0)
     8.204948872098814
```

Fig 9: Computing the Collective Diversity

The comparison of these two diversity measures which are used for recommendations are represented in a graph shown in Figure 10.The x-axis indicates the names of the diversities measured and the y-axis represents the measure of the diversity. From the graph shown in figure 10 it indicates that the collective diversity is more than the mean diversity obtained. Which means in summary the different user, cuisine and restaurant objects have a collective affinity between them which is hidden and can be represented using collective diversity.

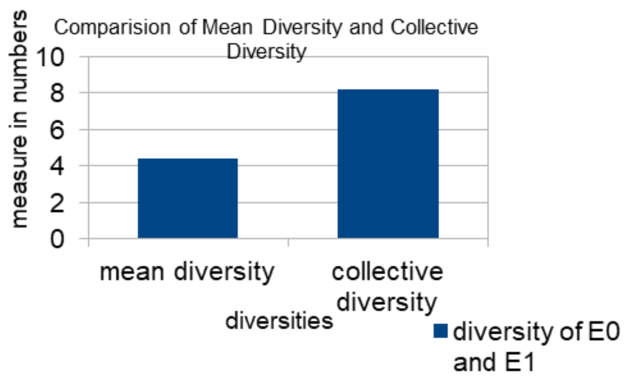

Fig 10: A comparison of mean and Collective Diversity between E0 and E1.

5. Conclusion

HINPy is a python workbench used on Heterogeneous Information Networks and is best suitable for the analysis of recommender systems. In conclusion, recommender systems play a pivotal role in enhancing user experience by efficiently filtering and presenting relevant content in a vast sea of information. Through the analysis of user preferences, behavior, and item characteristics, these systems employ various algorithms to generate personalized recommendations, ultimately improving user satisfaction and engagement across diverse platforms.HINPy is also used as a python library for applications in Artificial Intelligence, Machine Learning.

In addition to the theoretical exploration of recommender systems, this paper introduces practical and experimental functionalities aimed at implementing and evaluating recommendations using classic recommender system approaches. The inclusion of these experimental features enriches the scope of the research, allowing for hands-on applications and assessments.

The paper extends its contribution by incorporating functionalities that enable the evaluation of recommendations through classic diversity metrics. This practical aspect not only validates the theoretical foundations discussed earlier but also provides a tangible means of measuring the effectiveness and diversity of the generated recommendations.

An innovative element of the paper lies in the experimental extraction of time-series data related to apportionment and aggregations, specifically focusing on diversity measures. This temporal dimension adds a dynamic layer to the evaluation process, capturing how recommendation diversity evolves over time. This experimental approach can offer valuable insights into the temporal aspects of recommender systems, shedding light on their adaptability and responsiveness to changing user preferences.

By integrating both theoretical discussions and experimental functionalities, this paper not only advances the understanding of recommender systems in theory but also bridges the gap between theory and practical implementation. This comprehensive approach contributes to the broader field of recommendation systems research, fostering a deeper understanding of their dynamics and capabilities.

References

[1] Sun, Yizhou, and Jiawei Han. "Mining heterogeneous information networks: a structural analysis approach." Acm Sigkdd Explorations Newsletter 14.2 (2013): 20-28.

[2] Liu, Jiawei, et al. "A survey on heterogeneous information network based recommender systems: Concepts, methods, applications and resources."AI Open 3 (2022): 40-57.

[3] Sun, Yizhou & Han, Jiawei & Aggarwal, Charu & Chawla, Nitesh. (2012). When Will It Happen? — Relationship Prediction in Heterogeneous Information Networks. WSDM 2012 - Proceedings of the 5th ACM International Conference on Web Search and Data Mining. 663-672. 10.1145/2124295.2124373.

[4] Yizhou Sun, Yintao Yu, Jiawei Han, "Ranking-based clustering of heterogeneous information networks with star network schema ",Proceedings of the 15th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (2009), pp. 797-806.

[5] Yizhou Sun, Jiawei Han, Xifeng Yan, S Yu Philip, Tianyi Wu,"Pathsim: meta path-based top-k similarity search in heterogeneous information networks",Proceedings of the VLDB Endowment, 4 (2011), pp. 992-1003.

[6] Lao Ni, William W. Cohen,"Relational retrieval using a combination of path-constrained random walks",Mach. Learn., 81 (1) (2010), pp. 53-67

[7] Jing Zhang, Jie Tang, Bangyong Liang, Zi Yang, Sijie Wang, Jingjing Zuo, Juanzi Li,"Recommendation over a heterogeneous social network ",2008 the Ninth International Conference on Web-Age Information Management, IEEE (2008), pp. 309-316.

[8] Pierre De Handschutter, Nicolas Gillis, Xavier Siebert,"A survey on deep matrix factorizations",Computer Science Review,2021,Volume 42.

[9] Chuan Shi, Xiangnan Kong, Yue Huang, S Yu Philip, Bin Wu"Hetesim: a general framework for relevance measure in heterogeneous networks",IEEE Trans. Knowl. Data Eng., 26 (10) (2014), pp. 2479-2492.

[10] Xiaotian Han, Chuan Shi, Senzhang Wang, S Yu Philip, Li Song "Aspect-level deep collaborative filtering via heterogeneous information networks",IJCAI (2018), pp. 3393-3399

[11] Huan Zhao, Yingqi Zhou, Yangqiu Song, Dik Lun Lee, "Motif enhanced recommendation over heterogeneous information network",Proceedings of the 28th ACM International Conference on Information and Knowledge Management (2019), pp. 2189-2192

[12] https://archive.ics.uci.edu/dataset/232/restaurant+consumer+data

[13] Sadhana Kodali, Madhavi Dabbiru, B Thirumala Rao ,"A Cuisine Based Recommender System Using k-NN And Mapreduce Approach",International Journal of Innovative Technology and Exploring Engineering (IJITEE) , Volume-8 Issue-7 May, 2019