DEEP LEARNING BASED APPROACH FOR PERSONALIZED RECOMMENDATIONS

By

Junaid Tariq

A thesis submitted to the faculty of Electrical Engineering ,
Military College of Signals, National University of Sciences and Technology,
Islamabad, Pakistan, in partial fulfillment of the requirements for the degree of MS in
Electrical (Telecom) Engineering

May 2021

ABSTRACT

Knowledge-graphs are the most effective type of things used by the google to improve and provide the best search result to the user. Knowledge-graphs contain the well-structured information about the users based on entity to user relation. Most of the researcher used the knowledge-graphs to cope with the cold-start and the sparsity-based problems due to its effectiveness. However the techniques which are already proposed mainly rely on manual feature engineering and did not allow the end-to-end training. Similarly, most of the techniques based on homogeneous based KG's and very few are based on heterogeneous based KG's. So, there is a need of such technique which not only solve these problems but also help system and model to improve its performance which we will discuss later in depth.

Here we propose the deep learning based approach for personalized recommendation with label propagation algorithm which computes the user-item embedding for the particular user which is based on Graph SAGE and trained in the way of GNN like images.Moreover for the experiments In order to know the desired probability we first we construct the knowledge-graph after taking the text file from Microsoft satori and coupled it with the data which is being used to construct the more generic.

Knowledge-graph for taking the specific users preferences we first take the user-item embedding for the specific user after applying the scoring function after that we used the labeling algorithm to provide the better labels for the data which is initially unlabeled to provide batter labels so it can be batter represent the neighborhood labels as compared to the baselines. We introduced the label propagation algorithm a semi-supervised learning algorithm which is used for the efficient labelling of the unlabeled data points. By the efficient labelling after we got entity features and results from the GNN model we used it for the efficient labelling to make perfect assumptions. We prove and show our results on the two publicly available data sets namely LAST.FM and movie-lens data sets along with the results we showed our model effectiveness and proved the performance of our method is best as compared to baselines.

DEDICATION

This thesis is dedicated to

MY FAMILY, FRIENDS AND TEACHERS

for their love, endless support and encouragement

ACKNOWLEDGEMENTS

I am grateful to God Almighty who has bestowed me with the strength and the passion to accomplish this thesis and I am thankful to Him for His mercy and benevolence. Without his consent I could not have indulged myself in this task.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACRONYMS

Recommender Systems	RS
Collaborative knowledge-base Embeddings	CKE
Deep knowledge aware Network	DKN
Personalized Entity Recommendation	PER
Graph Neural Network	GNN
Label Propagation Algorithm	LPA
Knowledge-graph	KG
Label smoothness	LS
Singular Value Decomposition	SVD

Chapter 1

INTRODUCTION

With the rise of technology and the advancement of the AI in the new era has revolutionized the Human life and it effects on each and every part of life from the daily life to online marketing we are in the recommendation engines.Let's suppose we visit the shop of grocery or cloth shop the shopkeeper will tell us or show us the things according to our previous choices [1].Similarly if we see the online systems like Amazon,YouTube,Netflix and google and similar types of online systems they recommends the things to the users based on their preference and the taste.There are several types of recommendation systems that currently being used in the media and the industry.But we should also keep in mind that the pros and cons of the recommendation systems as the recommendation systems are recommending the things to the specific types of users according their taste and preference [1].But what if the new user having no previous record is present then in this case what will be the solution in that case most of the times the systems and the persons take the side information and also used the content based recommendation systems but users and the systems most of times prefer to use the side information like as knowledge graphs and the some sort of such type of information which can assist the new users according to their preferences etc.As we discussed earlier there are several types of the recommendation systems but there are six main types of recommendation systems which are currently practised in the media and industry which are discussed below in detail [1].

1.1 Thesis Statement and Objectives

Existing KG Based recommendation methods are mostly homogeneous(KG). Few heterogeneous Knowledge graph based recommendation systems are proposed in literature but they require extra meta paths and manually designed features which is very tedious for larger recommendations.

Thesis Objectives The research aims to achieve the following goals:

- To purpose a technique based on heterogeneous KGS as a hybrid KG-aware recommendation method for the personalized recommendation

- To compare the proposed model with the state of art techniques to show its effectiveness.

1.2 Contributions

The main contributions of this research is that here we checked the behaviour of heterogeneous Knowledge graphs with GNN and LPA.We can also use graph sage but having said that for it we need the larger datasets but as we have two different datasets having different lengths so we preferably used the GNN and for perfect labelling and the identification of each and every user(node) we used the LPA which is semi supervised machine learning algorithm.So we can say that we have two main contributions here.

- **GNN** Checking the behaviour of GNN on heterogeneous knowledge-graphs.
- **LPA** For labelling and prefect classification we used the LPA (Label propagation algorithm).

1.3 Content based Recommender Systems

In content based Recommender systems the recommendations are made based on the content of the particular product to the specific user depending on the needs and preferences of the particular user and targeted public and community [2].

1.4 Collaborative Filtering

Collaborative filtering methods recommends the things to the users by collecting the many users preferences and the tastes based on preference.The basic theme of recommendation in the collaborative filtering is that it collects the preferences of the community and based on their preferences it recommends the item to the users [42].

1.5 Utility Based Recommender Systems

Utility based recommendation systems recommend the item to the users based on the utility and consumption of the specific item with respect to the specific user.The utility of the item also depends on the community of the certain place that how much a community uses the certain item.Based usage the item recommends to the users [3].

1.6 Knowledge Based Recommender Systems

A Recommender system is knowledge-based when it makes recommendations based not on a user's rating history but on specific queries made by the user [4].

1.7 Cross-Breed based Recommendation systems(Hybrid Recommendation systems

When we combine the two or more recommendation systems then the new type of recommendation system forms which is called as the hybrid based Recommender systems.Basically hybrid based Recommender systems form by the combination of two or more Recommender systems [2].

1.8 Demographic Based Recommender Systems

Based on the demographics such types of recommendation systems make recommendations to the specific users and recommend the things to the users.Due to its simplicity and the straight forwardness the many industries took this approach and implemented it practically.Before the implementation of the demographic based recommendation systems a proper survey is made in the market and the certain place where the such type of the recommendation systems will be deployed.The most important thing in the demographic based recommendation system is that it requires the people to people contact and their surveys and impact of their behaviour on the certain type of things in the market. The most important advantage of the Demographic based recommendation system is that it does not require any kind of back end or precious knowledge to make the recommendations [5]

1.9 Techniques Used

We used the GNN on the heterogeneous-KGs to check the performance of GNN on heterogeneous KGs that how much they giving good results in such case. For the efficient labelling and perfect classification we also used the label propagation algorithm which gave the prefect labels to the classes of users based on their preference and choices. [18].We proved our techniques with two publicly available datasets like FM and Movielens 20M through them we got perfect results as compared to the state of the art baselines.

1.10 Thesis Outline

This thesis is divided into five chapters:

- Chapter 1: This chapter contains introduction and objectives. It also contains the contributions we have made in this thesis report.

- Chapter 2: In this chapter, review of literature and background is given along with brief description of existing technique and quantitative measures used in this report.
- Chapter 3: In this chapter we completely described the used approach along with the Label Propagation Algorithm that we used in our technique.
- Chapter 4: Here we showed the effectiveness of our proposed technique on the publicly available datasets and proved that the LPA with GNN is much better as compared to the baselines in terms of Accuracy Recall and Top@k recommendations.
- Chapter 5: Here we concluded our research and also along with that we showed some future directions like Label smoothing and Graph SAGE with LPA.

Chapter 2

PRELIMINARIES

2.1 Recommender Systems

The recommendation engines are the algorithms which recommends the precise item to the user supported his interest and profile on which he paid attention before for this the eye based collaborative filtering which recommends the sole specific a part of the item was proposed [49].Like wise on the idea of the implicit feedback the Bayesian personalized recommendation divides the users preferences into the three parts for this we should always now the implicit and the specific feedback in explicit feedback if a user just watch an item and didn't gave any feedback there than it'll dwell the category of the specific feedback which is explicit feedback.If a user watched and gave some rating there than it's called the implicit feedback.In case a user neither watch nor gave any feedback before the BPR it had been treated similar with the specific or negative sign. In BPR the item on which user gave the some rating or likeness is represented with the positive sign and therefore the during which he didn't gave feedback and just watched the item is represented by the negative sign similarly for that item during which he didn't watch and neither gave any feedback is represented by the blank. space [48]. These are the some attention based recommendation methods.

2.2 Multitask Feature learning for Knowledge-graph enhanced learning

In MKR the which uses the KG embeddings for the advice purpose. MKR is especially composed of the cross and compress units which automatically learnt the upper order interactions between users and items and therefore the entities relation within the KGs. However this system don't allow end-to-end training [31]. The proposed framework of the MKR is shown in figure below with the cross and compress units.

2.2.1 Work with cross and compress units

For interoperability models of objects and equipment, cross and compress devices were created in MKR frames. For point v and one of the intersections, the dd-pair interactions between their latent characteristics have flRdand elRd formed by layer 1: where ClRdd is the

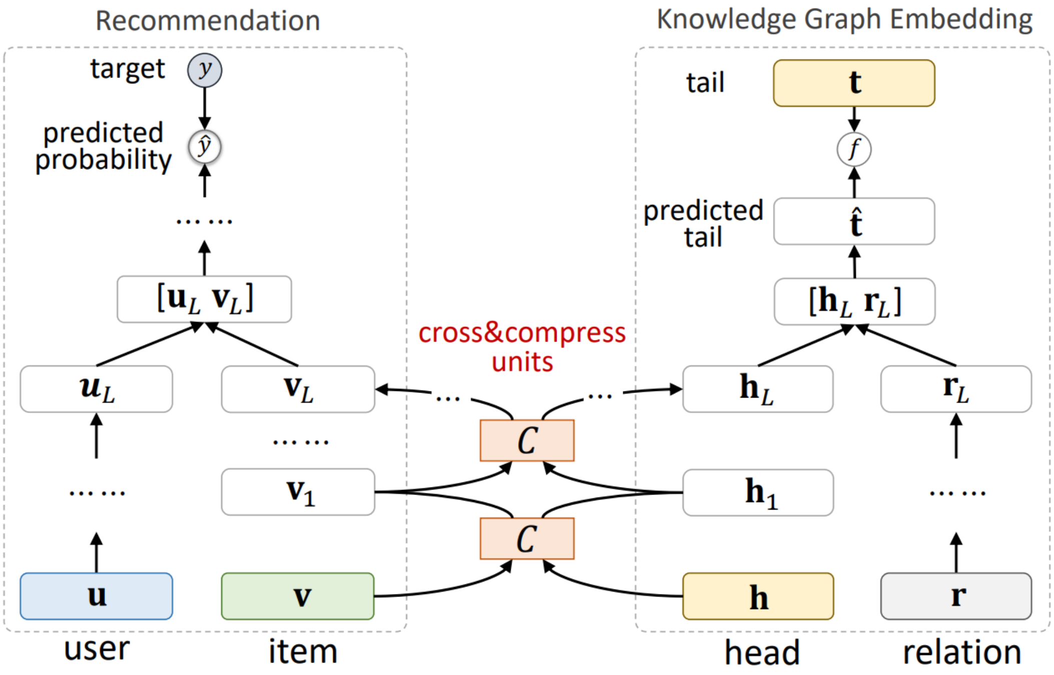


Figure 2.1: MKR Framework

variable of the work matrix for layer l, and d is the size of the hidden layers. This is called the rotation function, because an alternating function is v (i) le (j) l, (i, j) 1, ..., d2 between v and the intersecting columns e are clearly shown in the cross section [31]. The working environment system is divided into both vertical and horizontal for the disadvantages of uniformity. For simplicity, the cross and compression units are shown in figure where v and e do not normally output in this process. Using cross-linking tools, you can adjust the weight of the transmitted data and learn the accuracy of the two functions. It should be noted that cross-compression units should only be available in low MKR. In colleges, jobs are often shifted from broad to network-based, and changes can be reduced at higher levels with increased productivity [31]. Therefore, that of the high-grade layer may have a low transmission risk, especially for heterogeneous operations in MKR [31]. In low MKR layers, sentence functions were mixed with user functions, and attributes of objects were mixed with similar functions. Mixed tasks are not suitable for division when they do not require mutual understanding [31].

2.3 Collaborative filtering knowledge base Embedding

In this paper, we explore heterogeneous data in a way that improves the structure of the consensus. First, by applying practical knowledge, our process is designed to present the representation of information from the design process, the script and the visual content, in chronological order. To be specific, a system of interconnected tiles, called TransR,is designed to present a model representation of the content by looking at the inconsistencies between the two nodes and relationships.Using stacked denoising auto-encoders and stacked auto-encoders,both types of in-depth instruction,to eliminate representation of topics,visible representation, respectively. Finally, the final employment agreement, called the CKE,to learn about the representatives of the joint ventures, along with the details on a regular basis represented by practical knowledge. Integrating filtering with cognitive input combines CF with cognitive, textual and visual cues in one sentence, but CGU module in TransR is better for graphical applications such as CG completion and the prediction linking.The CF-based module and CKE module in the Bayesian role make the CG supervisory for specific recommendations [32].

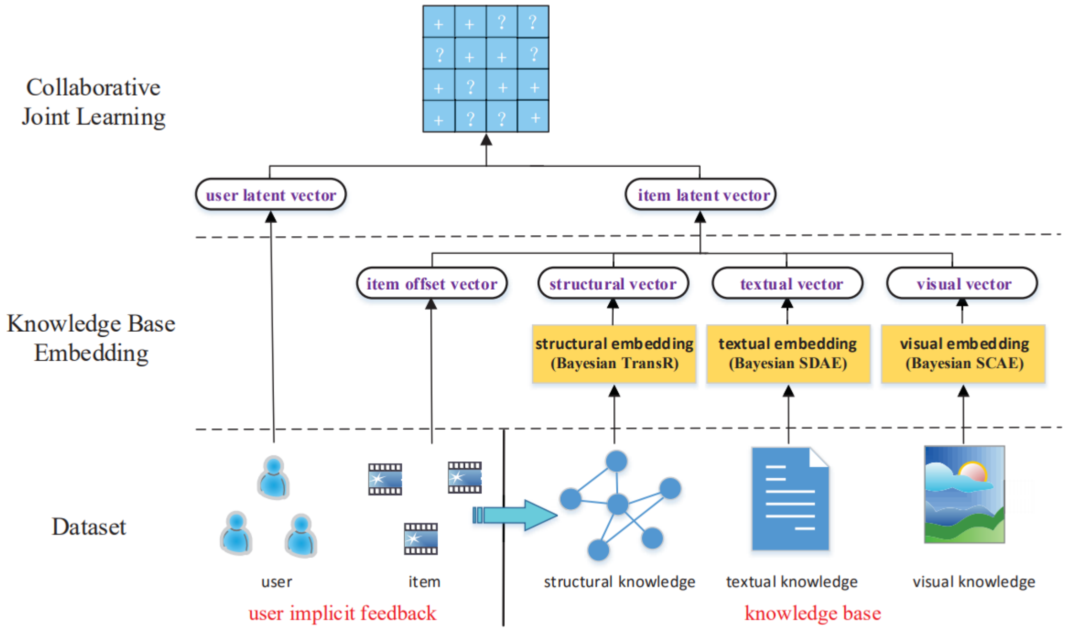


Figure 2.2: CKE Framework

2.4 RipllNet Propagating User Preferences on The knowledge Graph for Recommender Systems

Like Ripple Net, it falls into the category of integrated technology as agreed with imaging,its a memory network,just like the true ripples of water,Ripple Net supports the transmission of the user prefers to skip the dataset devices by itself and add the user equipment needed to connect in the familiar image.The framework of the ripple net is shown within the figure below [50]. The main issue which relates with the Ripple net is that here during this technique the importance of the knowledge is weakly categorized, and therefore the quadratic equation isn't well explaining the connection of users and the items [50]

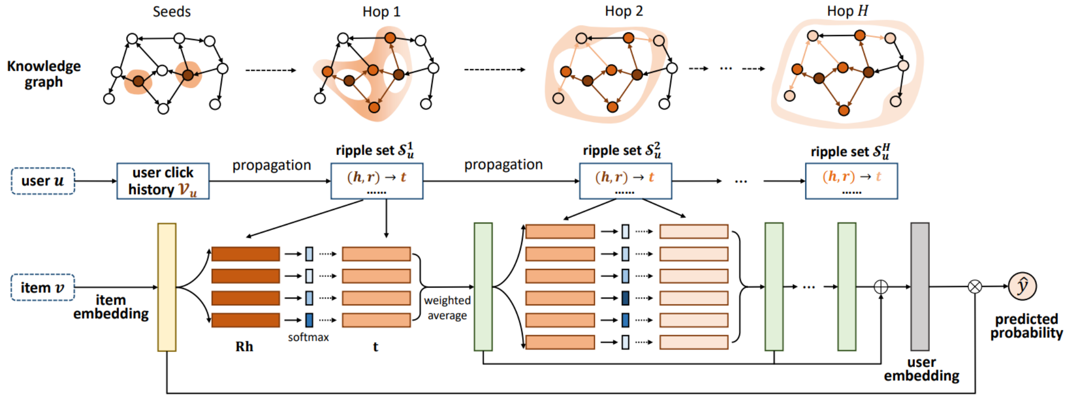


Figure 2.3: RippleNet Framework

2.5 DKN Deep Knowledge-aware Network for News Recommendation

DKN (a well-known network),which uses CNN and communicates with the CNN framework to discuss storytelling. The DKN framework is shown below.But here in DKN we need pre-configuration of the device, and it rarely pages apart from the script [30].

2.6 Personalized entity Recommendation a heterogeneous information network Approach

Similarly, PER, an integrated information sharing system, falls into the category of technology agreeing with familiar images. It monitors KG because of the heterogeneity and by requires the meta paths representing the connection between the element and the user. With a wide range of variations, PER is the appropriate process for systems like heterogeneous KG support. Its framework is shown in the figure below. The main problem with PER is that it requires additional modeling and design for the project. Similar to the group of factorization machine with lasso (FMG) [13].As this is the heterogeneous KGs but they are only suitable for the smaller recommendations and for the larger recommendations they requiring the extra meta paths and additionally designed features which limits its performance in generic recommendation scenarios [35].

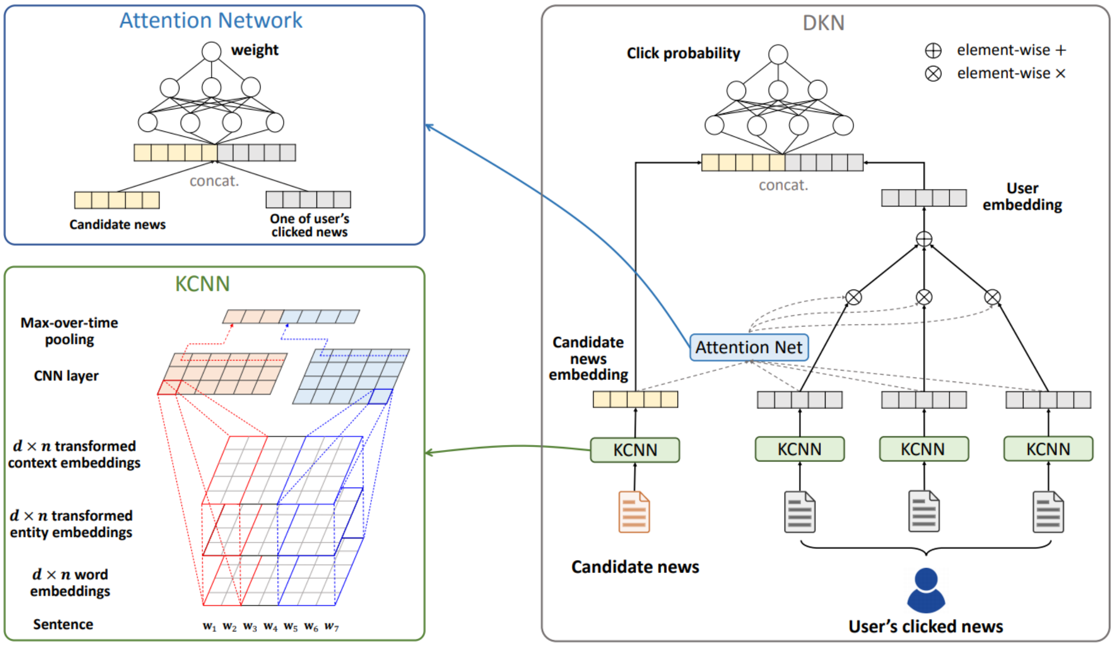


Figure 2.4: DKN Framework

Knowledge-Graphs

Prior to moving forward towards the GNN and its previous work we should know about the knowledge-graphs which is an integral part of our research because we used them as a side information to cope with the cold start and sparsity based problems to check the performance of GNN on heterogeneous KGs.So its necessary to have full knowledge about KGS in detail. Visual representation presents a model that connects organizations - products, events or concepts.Graphs provide information in context through links and sequential texts,and it provides a framework for data mixing,joining,analysis and sharing [14].

2.6.1 Homogeneous Knowledge-Graphs

Such types of Knowledge-graphs which are dealing with only one type of Users or just one class of users is know as homogeneous. knowledge-graphs [14] [42].

2.6.2 Heterogeneous Knowledge Graphs

Such knowledge-graphs which are coping with different types of classes of users and people and made the search easier for them is called as the heterogeneous knowledge-graphs.Usually all the knowledge-graphs are heterogeneous [14] [42]. In the next section we will see in detail the different types of knowledge-graphs. Knowledge Graphs (KG) represent the identification of descriptions of sites - products and events or uncertain concepts

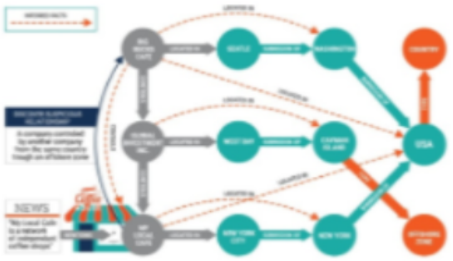


Figure 2.5: Knowledge-Graph Depiction

(e.g. Information).

- Descriptions contain semantics that can be applied to humans and computers to achieve their benefits and accomplishments.

- The device supports at least one sharing as a network, where a device represents a part of the network of devices interacting with it, and provides the results words for his translation [42].

2.7 Key Points

The key points of the Knowledge-graphs are described below in detail.

- **Database** Because most data is analyzed by standard queries.

- **Graph** It is identified as a network of two data formats.

- **Knowledge Bases** They have semantics that can be used to interpret information and provide new information.

- The diagram in RDF provides the most basic information for data sharing, joining, linking and retrieval.

- **Presentation** Templates in semantic web - RDF (S) and OWL - allow Apple to print between different files and concepts: data schema, taxonomy and terms. RDF * continues to simplify root templates and other metadata processes [14].

- **Performance** Each special instruction has been well thought out and proven to ensure effective control measures with thousands of facts and characteristics [14].

- **Interactions**Has many specialized features for data distribution, access (SPARQL endpoint protocol), administration (SPARQL setup store) and organization functionality. The use of international special characters leads to mixed data and publicity [14].

- **Designing**All of the above have been modeled by the W3C Community Standards to ensure that the needs of different workers are met - by entire engineers to the data management staff and working group [14].

2.8 Ontology and Semantics

Ontologies are the main part of the knowledge-graphs and they have the basics of the knowledge-graphs because they are the basic building blocks of the knowledge-graphs and interconnected with each other. Here the user may be another person or software program who wants to interpret the terms reliably and accurately. Ontologies ensure that information and its contents are shared [14]. While semantics concepts are often expressed and interpreted in terms of knowledge charts, there are several representations and models:

- **Processes** The description of the device usually consists of a classification of materials as well as a class process. For example, there may be individuals, organizations, and websites when it comes to marketing information. Individuals and organizations can have a standard superclass agent. Places often have different locations, for example, country, population, city, and so on.

- **Types Of Relationships** The connections of objects are usually marked with types that provide information about the nature of the connection, e.g. Friends, family members, competitors, and more. The parent-inverse is the relationship of the child, two special circumstances of the family, which can be the relationship. Or to mean that the extension area and the service center are changed [14].

- **Categories**An instrument often associated with categories that describe the significance of its core terms, such as ’The Greatest Expert’ or ’XIX Century Composers’. One book can be made for each of these categories at once: ’Books about Africa’, ’Bestseller’, ’Books by Italian Authors’, ’Books for Children’, etc [14].

- **Description** Most descriptions of ’human-friendly’ are provided in order to achieve the goal of developing and improving the research [14].

Figure 2.6: Big Knowledge-graph

What is not a knowledge-Graph? Not all RDF diagrams will be graphs of knowledge for example, a group of data sets, e.g. GDP data for countries, represented in RDF, not KG. The graphical representation of knowledge is often useful, but it is not necessary to capture knowledge content. It would be enough for an application to have only one string ’Italy’ related to the string ’GDP’ and more than ’1.95 trillion’ without having to mean which country posses.It is the connection and therefore the graph that KG makes, not the language that the data represents [14]. Not all miracles are a measure of knowledge.An important feature of KG is that the description of the device must be at least connected to the connection.The content of 1 device includes other components.This link is how the lights were created.(E.g. A is B. B is C. C is D. A is D).Technical knowledge does not have to be design and build,for example QA ”basic knowledge” some software also does not represent KG.It is possible to have a professional way of having the writing experience created in a non-graphical format,but using cut-outs in procedures,such as layers’ as that ’rules to aid analysis [14].

2.9 Examples of Big Knowledge-graphs

Google Map Google made this statement fashion with the release of its experience in 2012.However,there are a few details about its organization, services and size.There is also a small budget to use external maps of Google campaigns [14] [43] [44]. **DBpedia**This type of Knowledge-graph uses models of info boxes from Wikipedia to create large scripts of 4.58 items.(link https://wiki.dbpedia.org/about) and ontology of encyclopedic coverage of sites such as people,places,movies,books,organizations,species,diseases,etc [46]. **Geoname**In decline, users of Geonames data have access to 25 million units and locations [14]. **Wordnet** one of the best known lexical databases in English that provides translations and word processing available.Significantly improves performance for NLP and research [14] [43] [44] [45]. **FactForge** After many years of expertise in the broadcast media industry, Ontotext expands their knowledge of open source links and information about individuals, organizations and entities. It contains data from the KGs described above, also by specialists such as the Financial Industry [14] [43] [44].

Knowledge-graphs and RDF Charts

Over the years,we have moved away from different language with big data to Smart Data.With the value not known by the experience,this has led to the need for knowledge models that reflect our own understanding of information [14] [42] [43]. To make data smart, machines do not have to be bound by data without end schemes defined as 'a priori'.We need information stores that represent the 'real world' and therefore the tangled relationships that come with it.All of this needs to be erased from the machine and read the instructions as necessary to meet the needs of an automated system that fulfills and empowers ourselves [14] [45] [46]. RDF databases (also known as RDF triplestores),such as Ontotext's GraphDB, can assemble heterogeneous data from multiple sources and store thousands of facts about a single concept.The RDF structure is very robust [14] [45] [46]. As we have already seen, many libraries have links that can be accessed from sites such as DBpedia, GeoNames, Wikidata, and then they continue for a day. However, the main strength of the data line has become that when we triple our personal data to RDF and then connect our personal knowledge with open knowledge worldwide [14] [45] [46] [43]. Another important feature of RDF databases is the ultimate functionality, where new knowledge is often created from existing facts.Once these facts have emerged and been stored in an RDF database, our research can be comprehensive and provide new information for better understanding [14] [45] [46]. But if we want to make our data more powerful, we use text deletion to pull out important data from text-free streams and then add them to the facts in our root.

2.10 Uses of Knowledge-Graphs

Here we will see the different uses of Knowledge-graphs.

- Improves the visibility of business data easily.

- Build trust with a large audience wanting to try to find products / information [47].

- Provides trust between companies by analyzing reviews and metrics before users click on them [47].

2.11 Work Based on Graph Neural Networks

Burna et al. Interpret the broader range of affected neural networks through CNN, where images, video, and speech are represented, to higher realms, such as social networks,connections of the brain,or speech,through images.The presentation of models of CNNs in the framework of observing the aesthetic images that show the numbers required and subsequently optimized to make the site a fast filter of image.Significantly, the described process has a relatively uniform consistency and performance rather than that of the original CNNs,When applied to all structural configurations [16]. In these endeavors, according to the Neural Networks, Defferrard et al presented outstanding tools in visual and audio production benefiting because of their ability to achieve benefit from local translation changes of symbolic issues throughout their registry.In this form made possible by the publication from the CNNs for the campaigns mentioned for many of the original without any order from the interpreter in particular, two architectural structures, one based on the upper group of the collection, and the other supporting the problems of the Laplacian art [17].

Kipf liy proposed Scalable approach to monitor half of the incident-data that supports the performance of neural function that works directly on the image.Here the choice of this convolution architecture of an area determines closeness to the spectral graphical convolutions.This model scales linearly in the number of edges and learns the hidden representation layer that encodes both the local image structure and the characteristics of the nodes [18]. IN Share Description is a huge-scale recommendation engine that we have developed and implemented on Pinterest. PinSage Graph-Graph Convolutional Network (GCN) algorithm that mixes parallel and graphical lines to obtain inserts between nodes (e.g. points) containing two graphs text about node function. Compared with GCN's previous teaching approach,the improvement of performance standards facilitates the completion of the walk-through to the establishment and development of specialized training standards that rely on hard work and difficulty modeling to measure model strength and variability improvements.While developing a value for the process from the MapReduce model to use the learning model. We used PinSage on Pinterest and brought it to 7.5 billion examples on the graph with 3 billion nodes representing pins and boards and 18 billion rands. Consistent with offline numbers, user research, and A / B testing, PinSage develops improved data that is comparable to other methods of in-depth study and graphing [35].Monti et al in the Matrix modeling process explains that this is one of the most important aspects of the approval process. The complete matrix architecture combines different graphs of neural network graphs that show useful information about users and information, and repetitive neural networks that use graphical referrals. Here is the neural network system counts positive because it requires continuous numbers regardless of size [19]. The image of the known content retains the structural and social information of the group of objects or equipment. embracing the familiar graphs is a beautiful piece of information that can help improve visibility in use. Concepts, our way of counting users' specific agreements by prior to implementing activities to identify key knowledge issues for users provide. To do this, we transfer the knowledge data into the user of the physical device and then use the neural network to place the device.To provide better refinement,the label is assumed to be the same which means that the content next to the familiar knowledge charts will have the same form / score for users. The smoothness of the mark ensures constant passing of the blur weight and proves that it looks like a lot of labels on the drawing set.Data on visual performance can be correlated with the application to the size of the visual representation. Intellectual Property Development (KGCN), an end-to-end framework that monitors product relationships by speeding up their interaction of KG products.Identify both information about KG's work and information,the model of the neighbors for each unit in KG according to their location,then link the information in an environment when counting the representatives of a given group.The most common field is for more hops to model high-resolution data and capture users' long-term capitalization [21].Similarly, Schlichtkrull et al has also applied models with GCNs, although this is not applicable, but an agreement is required [22]. All of these methods discussed here were based on standard hemogenic knowledge; they do not focus on the GNN behavior of heterogeneity indicators. Some methods have been discussed before, which are heterogeneous, but it should be designed to create more functions and metaphats along with the end-to-end train of ideas that can work on heterogeneous knowledge with neural networks.

2.12 Semi Supervised Learning on Graphs

The main goal of the graph based semi supervised learning is to learn the proper labels to the unlabeled data so that it become easier for the model to predict the exact classes and made the best classification for this purpose the LPA we used which is a semi-supervised learning algorithm which initially operates on the some label points and computes the exact labels for the rest labels which are needed [23, 24, 25].Previous work makes assumptions on the graph through the smooth variations around the nodes of the graph. Having said that we can categorize the method in which edge weights are considered into two categorizes.1)The learnable edge weights in the fixed assignment of weights are considered as fixed so therefore fixed weights are given as input [23, 24, 25].Contrarily for the second assignment of weight the edge weights are not fixed so therefore means that the edge weights are parametric and learnable and they fixed or adjusted according to the requirement [23,24,25].If we talk about our technique in which we move with LPA which tackles the problem of edge weights and controls the over-fitting with the L2 regularization which not only give the best results but also reduce the training time efficiently.

2.13 Recommendation with Knowledge-graphs

Based on the previous experience and the literature we can divide the methods into three categories which are working with KGs.

2.13.1 Embedding Based Methods

Embedding based in which the KGE is preprocessed with the KG for recommendation embedding based methods are very effective type of methods used in recommendation with KG but such type of models which are based on KGE are best for in graph applications rather than recommendation [29, 30, 31, 32].

2.13.2 Path Based

Path Based which provides various pattrens of connections among items within the KG to supply additional guidance for the technique like PER heterogeneous and meta path heavily believe the additional meta path which is tough to optimize in practice and impossible to style the hand crafted meta paths in certain scenarios where entity and relations aren't within the same domain [34, 35, 36].

2.13.3 Hybrid Based Methods

The third one which is called as the hybrid is the combination of both of above techniques and learn the connection between user item in a very better way as compared to these methods our technique falls in the hybrid categories. [34].

2.14 Summary

Here in this chapter we discussed the relative research which is proposed before for with respect to the GNN and the LPA.

Chapter 3

DEEP LEARNING BASED APPROACH FOR PERSONALISED RECOMMENDATIONS

3.1 Graph Neural Networks

Geometry deep learning is explaining working of graphs on the neural networks.There are two important things in the graphs node and features so we should not mix nodes and features.So in our practical node is a person and the features are the characteristics of the person there are important matrices which we kept in our mind while working with GNN and these are Incident matrices,Adjacency matrices and Diagonal degree matrices [41].There are two types of graphs directed-graphs and undirected-graphs.

- Directed-graph:Such type of graphs in which the nodes are directed in only one direction is called as the directed graphs.e.g If a person follows an other person on a twitter it is not necessary that the person which is being followed also follow the second one so its graph will form as directed-graph [41].

- Undirected-graph:The undirected graphs are those graphs in which the direction of arrows is on both sides for example if a friend sends request to other friend on Facebook then the graph which will form in between them will be undirected one [41].

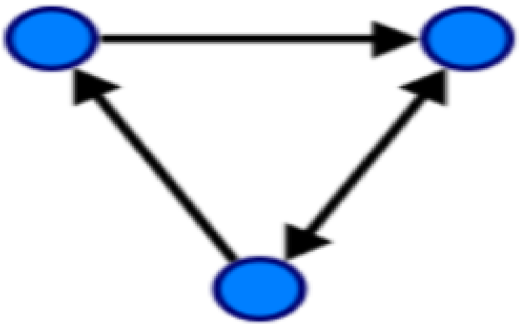


Figure 3.1: Directed and Undirected Graphs

3.1.1 Adjacency Matrix(A)

Adjacency matrix which is a square matrix usually used to represent the finite graphs.The components of the matrix are 0 and 1 depending upon whether the nodes of the graph are adjacent or not [41].

3.1.2 Diagonal Degree Matrix(D)

A diagonal degree matrix which contains information about the degree of every node that's the number of edges(connections) are connected with each node [41].

3.2 Laplacian Matrix(Graph Laplacian)

L = D-A Laplacian matrix which is also known as the Graph-laplacian.In Graph-Laplacian the Laplace Beltromi operator measures the smoothness of a vertix i.e how quickly it changes between adjacent matrix [41].In image processing first order differential of the image is $\nabla f = f(x+1, y) - f(x, y)$ which represents the smoothing of the images. where in A $\Delta$is equals to the Laplacian [41].

3.3 Why Convolution Fails on Graphs?

As if we see the images and texts these are the fixed grids because the neural networks which are designed these are only designed for the fixed grids so what should be done in the case where we have movement like GNN nodes.So to solve this problem we had two options for training of network like fixed grids.

- Adjacency matrix with the feature vectors.
- Second one is training the graphs like images using CNN

First approach is not very suitable because as in graphs the the nodes are continuously changing their position so the number of the neighbors also changes by the embedding of feature vectors in the adjacency matrix when the position of nodes changes then the number of neighbors of the target nodes also changes along with the feature nodes so this is not a effective [41].

When the images are trained on the network there are three main principles are required to follow in the CNN and by following these we also implemented this to our proposed technique as well and also did the training in the same way as in the CNN.The three main principles which must be followed during the training of CNN networks are.

- Locality
- Aggregation
- Composition (Function of a Function)

3.3.1 Locality

When we put the mask on an image basically we look the neighbors of the pixel and their effect on the center pixel which is done by locality principle so according to the locality principle whenever we are putting mask on the image we are basically looking at the neighbors of the pixel which is done in locality [41].

3.3.2 Aggregation

According to aggregation principle once we put the mask on the image we add all the neighborhood pixel and add it to the middle pixel which is our pixel of interest and this happens until the entire image completed so in CNN the locality operation is followed by the aggregation [41].

3.3.3 Composition Function of a Function

When we passing the features on CNN layers from one layer to a different layer than passing complex features to more complex on second layer is named as function of a function or composition.Composition consists of basic features of the CNN[40].So there are three problems associated with the Graphs when we use the CNN on them these are:

- Arbitrary size:They have complex topology.
- Graphs are Non-Euclidean
- No fixed Node Ordering

So when we used the CNN on images we usually used these three steps so in order to use these steps we have to train the graphs like images.So we will follow the locality,aggregation and function of a function like if we target on the node we used the neighborhood based aggregation and composition for this we will embed the graphs in the neural networks for this we have to look at the node embeddings [41].

3.4 Node Embedding

Node embeddings is the mapping of the nodes from the higher dimension to the lower dimensions in such a way that all the three principles of the CNN must be followed as like if we map the target node in lower dimension it should be like this that in lower dimension the distance between them as in higher dimensions should also remain same in lower dimensions and the by locality the neighbors of the nodes should be retained in lower dimensions and the aggregation in which the neighbouring nodes contribute in the target node.Similarly the complex features of the nodes passed in the neural networks so it is stacking of layers in the neural networks with the values of K.When we map the nodes from the higher dimensions to the lower dimensions than the size of the distance should remain there also.So by locality we have to check the immediate neighbors of the pixel let say c its immediate neighbors are A,B,D so locality means that how c is connected to the other immediate neighborhood pixels.And by aggregation means that how A,B,D contribute in C with their own weight matrixes.Lets say w1,w2,w3 that how these weights combine to contribute to C and the stacking of layers means that how can we perform more complex operations on graphs?By passing it into more complex layers [41].

Now after all discussion here we will first explain the construction of generic heterogeneous knowledge graph after that we will see that how we take the scoring function on the heterogeneous knowledge graph and we will construct weighted KG by the help of user item specific embeddings and develop the a KG for a person and in last the most important contribution is the model architecture with label propagation algorithm a semi-supervised algorithm which we will use for the efficient labelling of entities with specific user item interaction which shows the relation of a user with item.

Problem Formulation with Explanation

First, in the framework we had two main important things first the data and secondly knowledge graph.We had set of users U=u1,u2,u3,...and set of items I =i1,i2,i3,i4,... the main aim is here to predict the users potential interest in that item in which he didn't show any interest before for this we built the user item interaction matrix which is represented by

$$Yui = f(u, i) \tag{3.1}$$

which is defined in such a way that if a user interact with an item it will show as Y

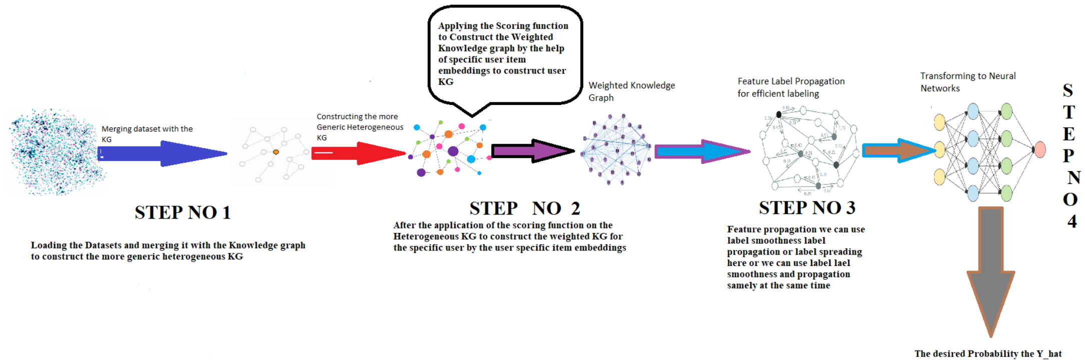
= 1 and if it didn't show any contact it is given as Y =0.So let's take on the knowledge graph now that how we combined it with the Data to make much better knowledge graph for the perfect recommendations according to the recommendation problems.As the KG has (h,r,t) here h ,r R and t which is denoting the head tail and the relation of the KG triples here and R showing the set of entities and the relations in the knowledge graph As we used two dataset so consider KG for movies to made the knowledge graph according to need we took itemindex2entityid,KG which is the Google Knowledge-graph and the movies ratings according to users feedback.Likewise for music we took itemindex2entity and the the userartistsinwhichtheuserIDartistID with weights is present which we put it with KG to prepare the KG according to requirement.We showed the KG in the figure 1 to give the idea of the knowledge graph.Lets consider the following example for KG triples as in Green Lantern is a film and the leading actor is the Ryan Reynolds which shows that the in KG the leading actor and Ryan Reynolds is hero in Green lantern.So in many recommendation I belongs to I to an entity e so we can say that the entity is consists of items and the I(I) as well as non items. After we constructed the required KG according to requirement our aim is to predict the following function which is shown in eq 1 with having user item interaction matrix.

$$\hat{Y}_{ui} = F(u, i/\hat{O}, Y, KG) \tag{3.2}$$

The framework of the proposed technique is shown in figure Now when we made the KG(prepared the KG according to our need) by the help of specific user item embeddings we took the scoring function of the KG which showed the users interest by weight on the specific items the scoring function is.

$$S_u r_e = g(u, r_e) \tag{3.3}$$

The scoring function helped us to see the users interest in depth for the particular item here as u and re are the feature vectors of the users u and relation type , similarly the g is the differential function which represents the inner product.Instinctively the $S_u r_e()$is designating the relations $r_e$with the user u.As an example if we see that a user shows more interest in directors of movies and other one shows more interest in the leading actor or genre of the


Applying the Scoring function to Construct the Weighted Knowledge graph by the help of specific user item embeddings to construct user KG
Merging dataset with the KG
Constructing the more Generic Heterogeneous KG
Weighted Knowledge Graph
Feature Label Propagation for efficient labeling
Transforming to Neural Networks
STEP NO 1
Loading the Datasets and merging it with the Knowledge graph to construct the more generic heterogeneous KG
STEP NO 2
After the application of the scoring function on the Heterogeneous KG to construct the weighted KG for the specific user by the user specific item embeddings
STEP NO 3
Feature propagation we can use label smoothness label propagation or label spreading here or we can use label lael smoothness and propagation samely at the same time
STEP NO 4
The desired Probability the Y_hat

Framwork of the proposed technique Deep learning based Approach for Personnalised Recommendations Consists of four steps 1)Constructing a heterogeneous KG 2)Applying scoring function to construct weighted KG 3) Applying Feature propgation(LPA,LS etc) and finally 4)Passing it to the neural networks to Get the Desired Probability Y_hat

Figure 3.2: Framework of Proposed Technique

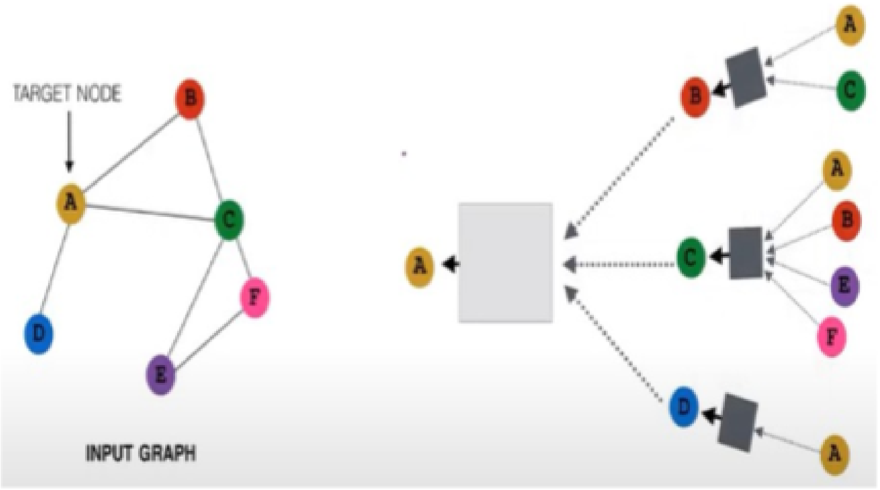


Figure 3.3: Feed Forward propagation of GNN

movie based on his preference [21].

As we transformed the KG into the lower dimension to check the users interest easily for this we take the adjacency matrix for user-item which is represented $A_u \in R^{|\in|x|\in|}$ where the (i , j) the entity $A_u^{i\,j} = S_u(r_{ei}, r_{ej})$and $r_(ei, ej)$ is the relationship between entities $e_i$and the e_j in the KG. If the $A_u^{i\,j} = 0$ than there is no connection in between the entities $e_i$and the e_j.The raw feature matrix entities are represented by the $ER^{|\in|X_0^d}$ here d_0 is raw entity feature.After we transform the KG into weighted KG we passed it to the GNN along with Label propagation for perfect classification during training using the collectively multiple feed-forward layers we can update the entity by the aggregation of neighboring entities.

$$H_{L+1} = \sigma(D_u^{-1/2} A_u D_u^{-1/2} H_L T_{uL}), L = 0, 1, 2, 3, ...L - 1 \tag{3.4}$$

the feed forward propagation of the neural networks is given as the in GNN is shown in the figure. Here we treat KG an undirected graph we had several candidates of like GCN and Graph SAGE we used the GCN as our base model and train it in supervised way we also had the Graph SAGE the difference between the Graph SAGE and the GNN based on GCNs is that it requires training in mini-batch for the larger graphs we had the movies data set which is initially coupled with the KG set however we also had the music dataset which is comparatively smaller as compared to movies keeping in mind we based our network architecture on the GCN and developed the GNN model which we trained in supervised way along with label propagation algorithm. For label propagation we will see in detail its working in our work and technique that how we used it for predicting labels of unlabeled data.Here it is important to mention that the GNN is almost same to graph SAGE the difference is in training.

In layers the matrix of hidden representation of entities is denoted by the $H_L$where H_0 equals to E.We represented the aggregator function by Au with respect to the adjacent neighboring entities or units.In this work during the whole process we set the $A_u = A_u + I$ which shows that for updating the old orientation is being considered.Here the diagonal degree matrix which is represented as D_u having entities representation $D_u^{i\,j} = \in_j A_u^{i\,j}$and $D^{-1/2}$ is used for keeping H_L(which is an entity representation matrix) stable along with normalizing of A_u.The layer wise specific trainable weight matrix is represented by the $W_L = R_1^{d+}{}_L^{d}{}^{+1}$ and represents the non linear activation function and the layers in the neural networks are represented by the L.

By the help of immediate neighbors and transformed mixture of single GNN layer the orientation of an entity or unit is computed in the knowledge-graph. Therefore,for checking the user's potential interest in depth the model can be extended to multiple layers.By merging the initial features along with their neighbors up to L hops away the final output which is the $H_L \in R^{|\in X}{}_L^{d}$ is calculated.Furthermore in the last the desired probability for a user u with the item I is calculated in which user's desired item's probability lies in which user has to show interest which is calculated by the formula $\hat{Y}_{ui} = f(u, i)$as we know that i_u(i.e. the nth row with having output H_L hop).Here it should be noted that the system is trained with on end-to-end way where the gradients of function flows from f(.) to g(.) for showing users interaction with items I.

3.5 Proposed Approach

3.5.1 Label Propagation with L2 Regularization

The main difference between the previous work and our work that is in the traditional GNN's the edge weights are fixed but here we set the weights$D_u^{-1/2}$,$A_u D_u^{-1/2}$ that are learnable which includes the possible feature vectors of users and items along parameter function g and they can be adjusted according to situation which demands the supervised signal and supervised way of training as the only signal of training comes from the user-item relation so there is a chance that the model will starts over-fitting.As we just now moving with the label propagation for handling this issue we introduced the L2 regularization in the more generic way and trained the model along with the LPA in the supervised fashion which gave the best results [25], [26], [27], [28], [29].

The citation along with regularization technique particularly L2 we obtain the best results we also got some advantages and benefits which are mentioned below. First, by the implementation of LPA with GNN the model's complexity decreased. Secondly along with LPA we got some great results in terms of accuracy and recall as compared to our baselines,Similarly the training time also decreased which shows the effectiveness of the LPA with GNN.

Label propagation which is our main contribution here is the semi supervised learning algorithm which assigns labels to the unlabeled data points for the efficient working of network which was initially unlabeled data points to represent the strong connection between the users and items.Consider an example of a movie lens dataset where some people like the movie due to title and some people like the genre or like the movie due to leading actor or director now the task is to predict the preference or taste of other people that either they are falling in these categories or not for this case to work in the LPA we made some assumptions on feature matrix and then transformed it into the label propagation for efficient labeling.For working concept we made an assumption an edge which is representing the connection of two nodes which showed the similarity so its mean that the entities in which connection occurs the users present their will show the same interest which was also being noticed during the implementation its mean that the relation entities with the user and items in graph which shared the same interest will lie in the same category.In the upcoming section we will see practically how it will work with the help of an example as it is shown in figure.

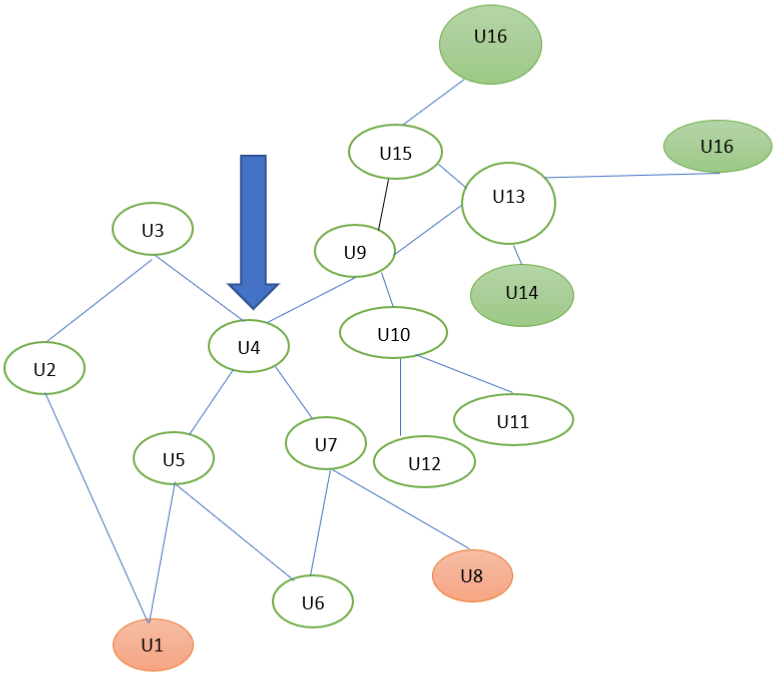


Figure 3.4: Label Propagation Algorithm

For the labels or class of the U4 we must move randomly,or we have to start the random walk from node 4 to predict that in which category it will lie. When there comes a labelled node, we will stop moving for this purpose we will see all possible routes going out from the U4 node.Here it is important to note that when we rich the node which has label then this node and this state will be called as the absorbing state. So, let’s start.

The possible walks which ends in the nodes representing the users which like Genre are given below from the figure 3. The walks are

1.U4→U9 →U15 →U16

2.U4→U9 →U13 →U14

3.U4→U9 →U13 →U15→U16

4.U4→U9 →U15 →U13 →U14

For the Orange node (Who like the director or Lead actor) the walks are which are shown in the figure 3

1.U4→U7→U8

2.U4→U7 →U6 →U5→U1

3.U4→U5 →U1

4.U4→U5 →U6 →U7 →U8

5.U4→U2 →U1

As we can see that the majority of the walks are ending on the orange one which is representing the class of people which likes the movies due to director or lead actor so its mean this user will also fall in this category based on this assumption and the label will be assigned to it.

The features which are obtained from the GNN for the efficient labeling of the nodes and entities we assign them the labels more accurately so that we can train it in the supervised way to know about the particular user's presence in more accurate way when trained this network with LPA after the GNN has applied. This is the basic idea and working of the LPA with GNN.The LPA representing with $R(A_u)$ Here first we will see the basic math formulation with the example and then we will see with the features the working of labeling classifier(propagation) with unified and supervised loss.

Consider the x_1 is the set of labelled nodes and the y_1 is the set of one hot labels of the labelled data. Consider there exists the 1, ..., c class labels where x_u represents the unlabeled vertices. Here the y_u is not known so it should be zero.

So, the random walks are given as

$$\mathbf{Y}_i[c] = \sum_{\mathbf{J} \in X_L} \mathbf{W}^t_{uij} \mathbf{y}_i[c] \tag{3.5}$$

Here the $y_i[c]$represents the probability of node $x_i \in \mathbf{X}_u$ having label c.Whereas the probability to move from the x_i node and terminate onx_j node in t foot steps. The number of steps can be defined to many steps(infinity).The matrix form of the equation will look like

$$\hat{\mathbf{Y}} = W^{t \to \infty} \mathbf{Y} \tag{3.6}$$

Here Y hat representing the target vector of which we must find the labels. The matrix of the probability having the classification of different nodes is represented by $W^{t \to \infty}$ Similarly, the vectors having labels is given by Y

ere in this equation we are interested in getting the labels for such type of users of which labels don't exists initially which we represented with the $\hat{Y_u}$

By computing the probabilistic transformation metric, we can calculate or find the labels of all the unlabeled nodes (here one thing should keep in mind node is representing a user and feature is its preferences or relations to items).By the help of the GNNs degree and the adjacency matrix we calculated the transformation(transition) matrix for all the nodes representing the user.

$$\mathbf{W_u} = \mathbf{D^{-1}_{u}A_u} \tag{3.7}$$

In equation we are normalizing the Adjacency matrix which is necessary for feature up gradation with respect to their weights.form here the newly obtained matrix is called s transisition matrix on which we performed labelling and got our desired results.
we divided the Before going further are mentioning here some notations which we will use.

- W_{LL} is representing the probability of the labelled to labelled nodes with in the transition matrix.
- W_{LU} is representing the probability of the labelled to unlabelled nodes with in the transition matrix.
- W_{UL} is representing the probability of the unlabelled to labelled nodes with in the transition matrix.
- W_{UU} is representing the probability of the unlabelled to unlabelled nodes with in the transition matrix.

Here it is important to note that the W_{LL} is representing the identity matrix and the W_{LU} is representing the zero matrix as due to the absorbing states we cannot move out from them. If we raise the power of W than it will get the larger values and started to saturate and the results which we will get are in steady state transition probabilities. As we can see that only two columns have the non-zero values and rest has zero values.The mathematical notation is as follows

.

$$\mathcal{W}_{\mathcal{U}} = \begin{bmatrix} W & 0 \\ W_{UL} & W_{UU} \end{bmatrix} = \begin{bmatrix} L & 0 \\ W_{UL} & W_{UU} \end{bmatrix} \tag{3.8}$$

$$\mathcal{W}_{\mathcal{U}} = \begin{bmatrix} L\&0 \\ W_{UL}\&W_{UU} \end{bmatrix} \tag{3.9}$$

$$\lim_{x\to\infty} W_u^t = \begin{bmatrix} L & 0 \\ W_{UL} & W_{UU} \end{bmatrix} \times \begin{bmatrix} L & 0 \\ W_{UL} & W_{UU} \end{bmatrix} \times \begin{bmatrix} L & 0 \\ W_{UL} & W_{UU} \end{bmatrix} \times ... \tag{3.10}$$

After multiplying these matrices we get

$$\begin{bmatrix} L.L + 0 + 0 + ...0 + 0 + 0... \\ W_{UL} + W_{UL}.W_{UU} + W_{UL}.W_{UU}^2 + ...W_{UU}.W_{UU}.W_{UU}.... \end{bmatrix} =$$

$$\begin{bmatrix} L & 0 \\ (\Sigma_t^{\infty}{=}_0 W_{UU}^t).W_{UL} & T_U^{\infty}{}_U \end{bmatrix} \tag{3.11}$$

where $\Sigma_t^{\infty}{=}_0 T_{UU}^t$representing the sum of geometric series x^n where determinant x is smaller than 1.

. We obtained the total sum of geometric series $(1-x)^{-1}$ As we self-multiplied W_{UU} the W_{UU} converged towards zero and it is always less than the 1 as I reached to 0.

Considering this graph as undirected graph which is shown in figure5 we have to find the labels of the rest of nodes with respect to the colored ones as it is undirected graph type so we can go in either direction when the absorbing state reaches the walk will end and we will trapped in between these nodes which are represented with the self-loops inside the graph.

so

$$\mathrm{W}_u = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0.33 & 0 & 0.33 & 0 & 0.33 & 0 & 0 \\ 0 & 0 & 0 & 0.5 & 0 & 0.5 & 0 \\ 0 & 0.33 & 0 & 0 & 0.33 & 0 & 0.33 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 \end{bmatrix}$$

As the node one for the user 1 which is representing the absorbing state has the probability 1 as it is absorbing so there is no way to go out from its so for the node to the probability for user to or node 2 will be zero.Similalry for others it will go like this and iteration will be so on. Contrarily from the node 4 which representing the user 4 we can go in node 1,3 and 5 so there is equal probability of moving that from the node 4 to 1,3 and 5 with the probability

0.33 for every node.Likewise, there is also equal probability of 0.5 to move from node 5 to nodes 6 and the 4.By the help of the GNNs degree and the adjacency matrix we calculated the transformation(transition) matrix or it can be taken as the normalization of A_u and we said new resultant matrix is called as the transformation matrix on which we performed the labeling and so we for this we splits the transformation matrix into 4 parts

$$\mathrm{W}_{u1} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$

$$\mathrm{W}_{u2} = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

$$\mathrm{W}_{u3} = \begin{bmatrix} 0 & 0 \\ 0 & 0.33 \\ 0 & 0 \\ 0 & 0.33 \\ 0 & 0 \end{bmatrix}$$

$$\mathrm{W}_{u4} = \begin{bmatrix} 0 & 1 & 0 & 0 & 0 \\ 0.33 & 0 & 0.33 & 0 & 0 \\ 0 & 0.5 & 0 & 0.5 & 0 \\ 0 & 0 & 0.33 & 0 & 0.33 \\ 0 & 0 & 0 & 1 & 0 \end{bmatrix}$$

with the help of equation 13

$$= \begin{bmatrix} R & O \\ 0.7318 & 0.2391 \\ 0.7318 & 0.2391 \\ 0.4854 & 0.4854 \\ 0.2391 & 0.7318 \\ 0.2391 & 0.7318 \end{bmatrix} \times 1/2 \begin{bmatrix} R & O \\ 1 & 0 \\ 0 & 1 \end{bmatrix} = \begin{bmatrix} R & O \\ 0.7318 & 0.2391 \\ 0.7318 & 0.2391 \\ 0.4854 & 0.4854 \\ 0.2391 & 0.7318 \\ 0.2391 & 0.7318 \end{bmatrix} sothelabelsweregivenlikethisasweshownr$$

3.6 Summary

Here we discussed the GNN and LPA in detail that how we trained the GNN and in what conditions the convolution fails o graphs so which technique we used to train the Graph on our network and along with that the proposed labelling scheme.

Chapter 4

EXPERIMENTAL RESULT AND ANALYSIS

For testing and verification of the proposed technique we took the two publicly available data sets and performed our experiments in three recommended data splits which we will explain here in detail along with different necessary parameters.

4.1 Datasets

In order to test and analysis of our proposed technique we used two publicly available datasets movie Lens 20M and music listening datasets along with that the core problems sparsity and cold start problems to assist the recommendations we used the Knowledge-graphs so that we can make our recommendations and classification best according to our requirements.

- **Movielens-20M:**Which is a widely used dataset we conducted our experiments on it by splitting the data in to the three different parts of recommended splits of data.The data is containing the ratings of 20M which is explicit and the total number of connections and entities in the corresponding entities are 102,569 and 499,474 edges with the 32 relation-types.

- **Last.FM:**Last.FM Data set which is based on the music listening of the two thousand users which contains the listenings of two thousands users.The dataset is divided into the three recommended data splits.Here the knowledge-graph entities are 9366 and edges which represents the connection between nodes are 15518 and contains the 60 relations between user and item set. The user item relations along with the entities and the number of connections among the users and items and entities relations are listed below in the table

4.1.1 Baselines

For comparing the results with state of art baselines we used the four baselines and compare our results with them and showed that our method outperform all of these.

Statistics	Movies	Last.FM
Users	138,159	1,872
Items	16,954	3,846
Interactions	13,501,622	42,346
Entities	102,569	9,366
Relations	32	60
KG triples	499,474	15,518

Table 4.1: Statistics of the user item interactions along with Kg triples representing entities and relations among user and items

SVD [51] which is based on collaborative filtering along with the inner product matrix for taking user item interactions for comparative analysis we used unprejudiced version of this model such as $y_{uv} = u^T v$.The parameters for the datasets and the proportions of the data are set as d=8,16,32,4,$\eta = 2e-2, 1e-2, 5e-4, 2e-4, forLast.Fm$we set the proportions and dimensions as d=32,16,8 and learning rates as 2e-2,5e-4,2e-4.

PER [35]Personalized entity recommendation belongs to the categories of the path based methods which works with the KG as the heterogeneous it extracts the meta paths and requires manually designed features for training and provides the additional guidance for the recommendations.The manually designed features are taken from the datasets of the Movielens 20M and the Last.FM. We set the proportions and the learning rate same as in SVD.

CKE [32] CKE which belongs to the embedding based systems in which knowledge-graph entities are coupled with the KG for recommendations and CKE divides the data into the structural visual and textual knowledge and represents the framework as a whole with the collaborative filtering method.The implementation of the CKE is done with the CF along with V,S and the textual knowledge.The proportions and the dimensions along with the learning rate which we set are 0.1,0.001,0.010.0001 and layers and dimensions are 8,16,32 etc.

RippleNet [32] which belongs to the hybrid based category method of the knowledge graph with the recommendation treats the KG as heterogeneous KG.Ripple net is also a memory network and which uses the users preference on the KG for recommendation.The proportions and the along with the learning rate are set as Lr = 2e-4,2e-5,4e-2,the LPA weight = 0.02,Dimension = 4,8,16,L= 2,4, batch size = 128,65536,256 and the L2 regularization =1e-4,2e-4 for both the datasets.

Moreover the aggregates we used labels and the sum aggregates along with the two number of iterations for each training set.

Techniques	Movielens 20M	Last.FM
PER [36]	0.832	0.633
CKE [34]	0.924	0.744
RIPPLE-NET [33]	0.960	0.770
SVD [51]	0.963	0.769
Proposed Approach	0.981	0.8168

Table 4.2: Accuracy wise comparison with the baselines

we also took the results of top@k recall and these results are given in table

Techniques	2kMovies	10kMovies	50kMovies	Movies100k
SVD [51]	0.036	0.124	0.277	0.401
PER [36]	0.022	0.077	0.160	0.243
CKE [34]	0.034	0.107	0.244	0.322
RippleNet [33]	0.045	0.130	0.278	0.447
Proposed Approach	0.045	0.160	0.3723	0.50

Table 4.3: Results of Top@k Movies Recalls

Techniques	2kLastFM	10kLastFM	50LastFmk	LastFm100k
SVD [51]	0.029	0.098	0.240	0.332
PER [36]	0.014	0.052	0.116	0.176
CKE [34]	0.023	0.070	0.180	0.296
RippleNet [33]	0.032	0.101	0.242	0.336
Proposed Approach	0.050	0.125	0.3150	0.4200

Table 4.4: Results of Top@k Last FM Recalls

The result of the top best recalls of the model wise comparison is given here in this table According to the recommended splits when we set the the training 60 percent 70 percent and the 80 percent we got the following results which are listed here in this table first we showed the full table then we will show the comparison wise with baselines.W took the datasets perportions into three formats like Training,evaluation and testing

Techniques	Movielens	Last.Fm
PER [36]	0.243	0.176
CKE [34]	0.322	0.296
RIPPLE-NET [33]	0.447	0.336
SVD [51]	0.401	0.332
Proposed Approach	0.50	0.4200

Table 4.5: Results of Top@k best Recalls

Now the comparison wise results of different combinations compared with the baselines.

Recommended Combinations of dataset of Used Approach	Movielens 20M	Last.Fm
60:20:20	0.978	0.806
70:15:15	0.9803	0.8119
80:10:10	0.9811	0.8168

Table 4.6: Results of Top@k Recalls

r	80percent	70percent	60percent
SVD [51]	0.955	0.913	0.882
PER [36]	0.828	0.802	0.821
RIPPLE-NET [33]	0.955	0.921	0.947
CKE [34]	0.921	0.898	0.916
Proposed Approach	0.9811	0.9803	0.9780

Table 4.7: Results of different ratios of datasets

4.1.2 Usefulness of LPA and L2 regularization

In order to learn about the usefulness of the proposed technique and the L2 regularization we conducted the experiments on the Last FM dataset.We took the Recalls of the model when we set the values of the L2 is set to 0 it performs well and the larger values of the l2 is less favourable because it can Leeds towards the poor results means the moderate values of the L2 regularization gives better results and perfect accuracies in CTR and Topk recommendation scenarios.In order to check the the efficiency of the LPA along with the L2 regularisation we fixed the proportions of the hidden layers and vary the values of the regularization between 0 to 5 and achieved the better results.The performance of LPA regularization with respect to the different dimensions are shown in the figure 5.

4.1.3 Results in cold and sparsity scenarios

The main aim of using the knowledge graphs in the recommendation is to cope with the cold start problems and the sparsity issues in order to cope with this we very the datasets and splits the data into the three recommended splits and these are training:evaluation:testing and splits are 80:10:10,70:15:15,60:20:20. The results of accuracies are shown in the table VI where we showed our results similarly in table V.These results showed that the proposed technique working well in the cold start and the sparsed scenarios [6].

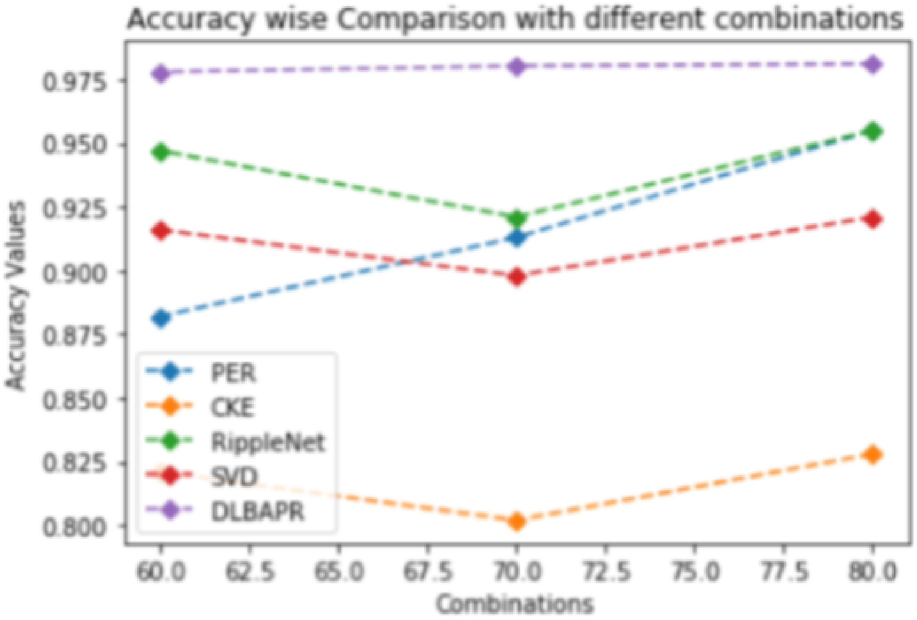


Figure 4.1: Recommended combinations and there results on Accuracy

. The dimensions of hidden layers which we used are very between 8 to 64 dimensions and the hidden layers 4 2 when we increases the d from a specific level it may decrease the systems performance and the over fitting may occur. Now here we will see the some visualization which are representing the results of CTR and the Top k visualization in figure 6 the accuracy of the models with movie dataset now in next figures we will show the results for accuracies on Music and than the recall wise results in topk recommendations.

The results for the different recommended combinations of the data sets with respect to model-wise comparison and the results in cold and sparsity cases are also given

The results for the different recommended combinations of the data sets with respect to the different model-wise comparison which also showing the results in cold and sparsity cases.

.

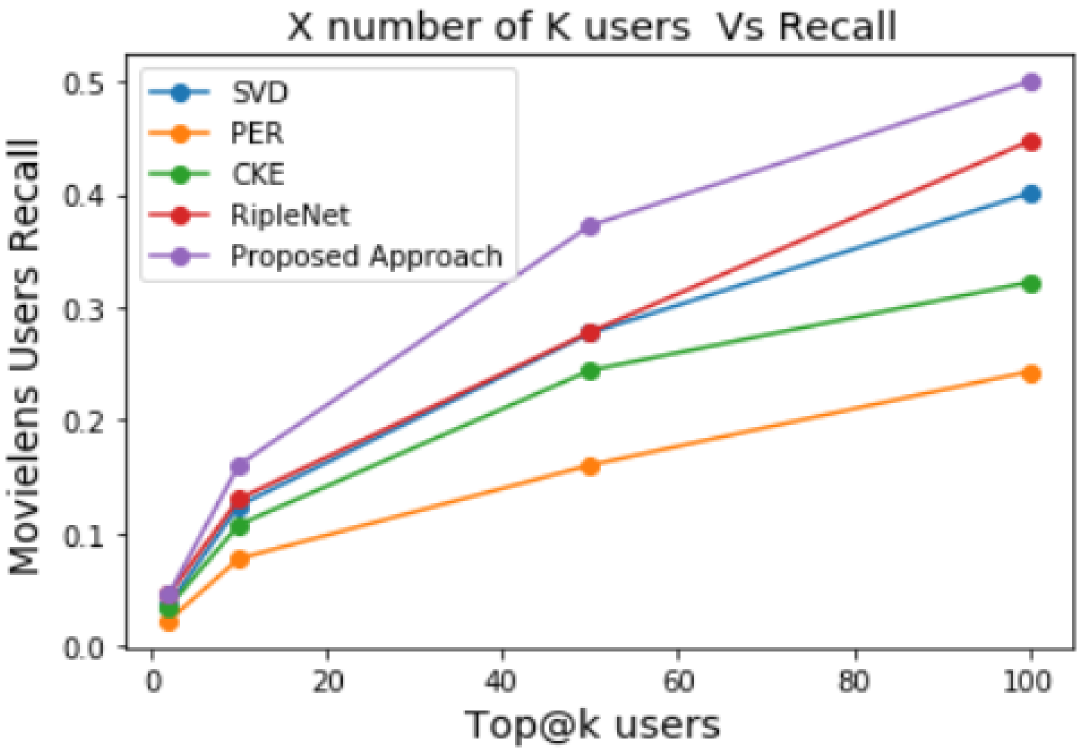


Figure 4.2: Movielens Recall

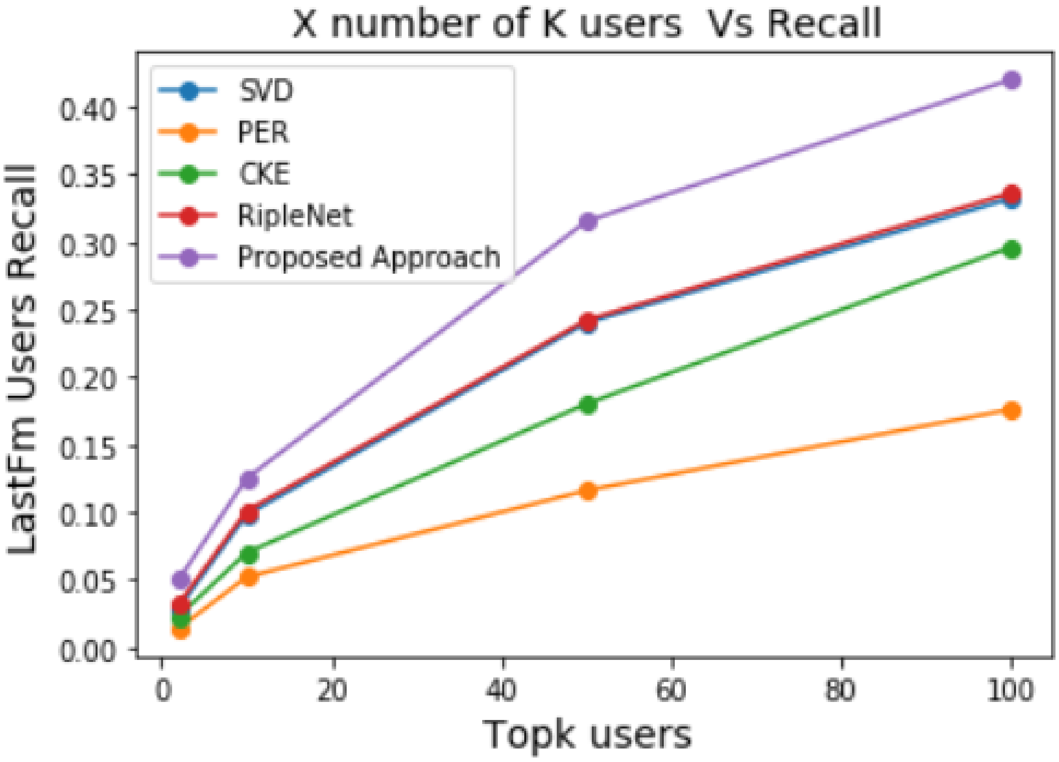


Figure 4.3: Music Users Recall

Chapter 5

CONCLUSION AND FUTURE WORK DIRECTIONS

As we know that the recommendation engines has the two main problems sparsity and cold start we proposed the Deep learning based approach for personalized recommendation were we used the LPA with the GNN and L2 regularization for the efficient labelling and the removal of the over fitting.We showed the results and checked the performance of the system with the proposed techniques in the cold start and sparsed scenarios and proved that the our method perform well as compared to the baselines.We also learnt that that how the edge weights are changes with the need and these are learnable.Beside these the methods which were presented before mainly was relying on homogeneous KGs with GNN here we checked the performance of GNN on heterogeneous KGs.

Here we want to elaborate specifically the LPA label propagation algorithm which is the game changer in our research which made the perfect recommendations and perfect labeling and categorized the number of users based on their features like some like movies due to lad actors and some due to Genre of the movie. Secondly the most important thing is Graph neural Networks.Here we faced two problems the changing direction of nodes and the labelling with movement we solved this by training the model like images which is fixed grid and having three important features like locality,Aggregation and Neighbourhood with embedding spaces that moving of the data from higher dimension to lower dimension in correct and exact way.

In future the relation between the label spreading and the label propagation would be a best choice and the Graph SAGE method as the architecture can be used here in place of GNN.Here we want to say that one there exists such probability like if both the classes have equal influence on the middle node then in which way it will go it can be solved with label spreading which is our future directions.

BIBLIOGRAPHY

[1] Ricci, F., Rokach, L. and Shapira, B., 2011. Introduction to recommender systems handbook. In Recommender systems handbook (pp. 1-35). Springer, Boston, MA.

[2] Pazzani, M.J. and Billsus, D., 2007. Content-based recommendation systems. In The adaptive web (pp. 325-341). Springer, Berlin, Heidelberg.

[3] Huang, S.L., 2011. Designing utility-based recommender systems for e-commerce: Evaluation of preference-elicitation methods. Electronic Commerce Research and Applications, 10(4), pp.398-407.

[4] Burke, R., 2000. Knowledge-based recommender systems. Encyclopedia of library and information systems, 69(Supplement 32), pp.175-186.

[5] Al-Shamri, M.Y.H., 2016. User profiling approaches for demographic recommender systems. Knowledge-Based Systems, 100, pp.175-187.

[6] Lika, B., Kolomvatsos, K. and Hadjiefthymiades, S., 2014. Facing the cold start problem in recommender systems. Expert Systems with Applications, 41(4), pp.2065-2073.

[7] Adomavicius, G. and Tuzhilin, A., 2005. Toward the next generation of recommender systems: A survey of the state-of-the-art and possible extensions. IEEE transactions on knowledge and data engineering, 17(6), pp.734-749.

[8] Guo, G., Zhang, J. and Thalmann, D., 2012, July. A simple but effective method to incorporate trusted neighbors in recommender systems. In International conference on user modeling, adaptation, and personalization (pp. 114-125). Springer, Berlin, Heidelberg.

[9] Jiang, Z. and Benbasat, I., 2004. Virtual product experience: Effects of visual and functional control of products on perceived diagnosticity and flow in electronic shopping. Journal of Management Information Systems, 21(3), pp.111-147.

[10] Massa, P. and Avesani, P., 2007, October. Trust-aware recommender systems. In Proceedings of the 2007 ACM conference on Recommender systems (pp. 17-24).

[11] Ziegler, C.N. and Golbeck, J., 2007. Investigating interactions of trust and interest similarity. Decision support systems, 43(2), pp.460-475.

[12] Shah, F., Bell, P. and Sukthankar, G., 2010, May. A destination recommendation system for virtual worlds. In Twenty-Third International FLAIRS Conference.

[13] Ziegler, C.N. and Golbeck, J., 2007. Investigating interactions of trust and interest similarity. Decision support systems, 43(2), pp.460-475.

[14] https://www.ontotext.com/knowledgehub/fundamentals/what-is-a-knowledge-graph/

[15] Improving the performance of recommender systems by alleviating the data sparsity and cold start problems G Guo - Twenty-Third International Joint Conference on . . . , 2013 - pdfs.semanticscholar.org

[16] Bruna, J., Zaremba, W., Szlam, A. and LeCun, Y., 2013. Spectral networks and locally connected networks on graphs. arXiv preprint arXiv:1312.6203.

[17] Michael Defferrard, M., Bresson, X. and Vandergheynst, P., 2016. Convolutional neural networks on graphs with fast localized spectral filtering. In Advances in neural information processing systems (pp. 3844-3852).

[18] Kipf, T.N. and Welling, M., 2016. Semi-supervised classification with graph convolutional networks. arXiv preprint arXiv:1609.02907.

[19] Kipf, T.N. and Welling, M., 2016. Semi-supervised classification with graph convolutional networks. arXiv preprint arXiv:1609.02907.

[20] Wu, Y., Liu, H. and Yang, Y., 2018. Graph Convolutional Matrix Completion for Bipartite Edge Prediction. In KDIR (pp. 49-58).

[21] Wang, H., Zhao, M., Xie, X., Li, W. and Guo, M., 2019, May. Knowledge graph convolutional networks for recommender systems. In The world wide web conference (pp. 3307-3313).

[22] Schlichtkrull, M., Kipf, T.N., Bloem, P., Van Den Berg, R., Titov, I. and Welling, M., 2018, June. Modeling relational data with graph convolutional networks. In European Semantic Web Conference (pp. 593-607). Springer, Cham.

[23] Baluja, S., Seth, R., Sivakumar, D., Jing, Y., Yagnik, J., Kumar, S., Ravichandran, D. and Aly, M., 2008, April. Video suggestion and discovery for youtube: taking random walks through the view graph. In Proceedings of the 17th international conference on World Wide Web (pp. 895-904).

[24] Kipf, T.N. and Welling, M., 2016. Semi-supervised classification with graph convolutional networks. arXiv preprint arXiv:1609.02907.

[25] Zhu, X., Ghahramani, Z. and Lafferty, J.D., 2003. Semi-supervised learning using gaussian fields and harmonic functions. In Proceedings of the 20th International conference on Machine learning (ICML-03) (pp. 912-919).

[26] Karasuyama, M. and Mamitsuka, H., 2013. Manifold-based similarity adaptation for label propagation. In Advances in neural information processing systems (pp. 1547-1555).

[27] Wang, F. and Zhang, C., 2007. Label propagation through linear neighborhoods. IEEE Transactions on Knowledge and Data Engineering, 20(1), pp.55-67.

[28] Johnson, R. and Zhang, T., 2007. On the effectiveness of Laplacian normalization for graph semi-supervised learning. Journal of Machine Learning Research, 8(Jul), pp.1489-1517.

[29] Huang, J., Zhao, W.X., Dou, H., Wen, J.R. and Chang, E.Y., 2018, June. Improving sequential recommendation with knowledge-enhanced memory networks. In The 41st International ACM SIGIR Conference on Research Development in Information Retrieval (pp. 505-514).

[30] Wang, H., Zhang, F., Xie, X. and Guo, M., 2018, April. DKN: Deep knowledge-aware network for news recommendation. In Proceedings of the 2018 world wide web conference (pp. 1835-1844).

[31] Hongwei Wang, Fuzheng Zhang, Miao Zhao, Wenjie Li, Xing Xie, and Minyi Guo. 2019. Multi-Task Feature Learning for Knowledge Graph Enhanced Recommendation. In Proceedings of the 2019 World Wide Web Conference on World Wide Web.

[32] Fuzheng Zhang, Nicholas Jing Yuan, Defu Lian, Xing Xie, and Wei-Ying Ma.2016. Collaborative knowledge base embedding for recommender systems. In Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. ACM, 353–362.

[33] Quan Wang, Zhendong Mao, Bin Wang, and Li Guo. 2017. Knowledge graph embedding: A survey of approaches and applications. IEEE Transactions on Knowledge and Data Engineering 29, 12 (2017), 2724–2743.

[34] Binbin Hu, Chuan Shi, Wayne Xin Zhao, and Philip S Yu. 2018. Leveraging metapath based context for top-n recommendation with a neural co-attention model.In Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery Data Mining. ACM, 1531–1540.

[35] Rex Ying, Ruining He, Kaifeng Chen, Pong Eksombatchai, William L Hamilton, and Jure Leskovec. 2018. Graph Convolutional Neural Networks for Web-Scale Recommender Systems. In Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. 974–983.

[36] Huan Zhao, Quanming Yao, Jianda Li, Yangqiu Song, and Dik Lun Lee. 2017. Metagraph based recommendation fusion over heterogeneous information networks. In Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. ACM, 635–644.

[37] Zhu Sun, Jie Yang, Jie Zhang, Alessandro Bozzon, Long-Kai Huang, and Chi Xu.2018. Recurrent knowledge graph embedding for effective recommendation. In Proceedings of the 12th ACM Conference on Recommender Systems. ACM, 297–305.

[38] Hongwei Wang, Fuzheng Zhang, Jialin Wang, Miao Zhao, Wenjie Li, Xing Xie, and Minyi Guo. 2019. Exploring High-Order User Preference on the Knowledge Graph for Recommender Systems. ACM Transactions on Information Systems (TOIS) 37, 3 (2019), 32.

[39] Hongwei Wang, Miao Zhao, Xing Xie, Wenjie Li, and Minyi Guo. 2019. Knowledge graph convolutional networks for recommender systems. In Proceedings of the 2019 World Wide Web Conference on World Wide Web.

[40] https://www.youtube.com/watch?v=1miz7yggcTg

[41] https://towardsdatascience.com/confusion-matrix-for-your-multi-class-machine-learning-model-ff9aa3bf7826

[42] Bonatti, Piero A. et al. “Knowledge Graphs: New Directions for Knowledge Representation on the Semantic Web (Dagstuhl Seminar 18371).” Dagstuhl Reports 8 (2018): 29–111.

[43] Paulheim, Heiko. “Knowledge graph refinement: A survey of approaches and evaluation methods.” Semantic Web 8 (2016): 489–508.

[44] Nickel, Maximilian et al. “A Review of Relational Machine Learning for Knowledge Graphs.” Proceedings of the IEEE 104 (2015): 11–33.

[45] Allen, J. and A. Frisch (1982). “What’s in a Semantic Network”. In: Proceedings of the 20th. Annual Meeting of ACL, Toronto, pp. 19–27.

[46] Shadbolt, Nigel et al. “The Semantic Web Revisited.” IEEE Intelligent Systems 21 (2006): 96–101.

[47] https://medium.com/analytics-vidhya/introduction-to-knowledge-graphs-and-their-applications-fb5b12da2a8b

[48] Rendle, S., Freudenthaler, C., Gantner, Z. and Schmidt-Thieme, L., 2012. BPR: Bayesian personalized ranking from implicit feedback. arXiv preprint arXiv:1205.2618.

[49] Fu, M., Qu, H., Moges, D. and Lu, L., 2018. Attention based collaborative filtering. Neurocomputing, 311, pp.88-98.

[50] Wang, H., Zhang, F., Wang, J., Zhao, M., Li, W., Xie, X. and Guo, M., 2018, October. Ripplenet: Propagating user preferences on the knowledge graph for recommender systems. In Proceedings of the 27th ACM International Conference on Information and Knowledge Management (pp. 417-426).

[51] Yehuda Koren. 2008. Factorization meets the neighborhood: a multifaceted collaborative filtering model. In Proceedings of the 14th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. ACM, 426–434.

[52] https://towardsdatascience.com/l1-and-l2-regularization-methods-ce25e7fc831c e nddocum[5]ent