JAYPEE INSTITUTE OF INFORMATION TECHNOLOGY

MINOR PROJECT-1 Synopsis

Job recommendation using Knowledge graphs

SEMESTER : 5th SEMESTER

BRANCH : INFORMATION TECHNOLOGY

BATCH : B-11

GROUP NO. : G63

COURSE CODE : 15B29CI591

STUDENT DETAILS: AMIT G. PATIL - 19104004

SANJOLI GOYAL - 19104007

MUSKAN JAIN - 19104010

Table Of Contents

Abstract

Dealing with the enormous amount of recruiting information on the Internet, a job seeker always spends hours to find useful ones. To reduce this laborious work, we design and implement a recommendation system for online job-hunting. Instead of using collaborative filtering(CF) algorithms we contrast on a Knowledge graph based approach to figure out more interrelations between candidates and job description.
Few works which utilize CF do not address the scalability challenges of real-world systems and the problem of cold-start. In this paper, we propose a scalable item-based recommendation system for online job recommendations. Our approach overcomes the major challenges of sparsity and scalability by leveraging a Knowledge graph of jobs connected by multi-edges representing various behavioral and contextual similarity signals. The short lived nature of the items (jobs) in the system and the rapid rate in which new users and jobs enter the system make the cold-start a serious problem hindering CF methods. We address this problem by harnessing the power of deep learning in addition to user behavior to serve hybrid recommendations.

INTRODUCTION

The increasing usage of the Internet has heightened the need for online job hunting.According to Jobsite Report 2014, 68% of online jobseekers are college graduates or post graduates. The key problem is that most job-hunting websites just display recruitment information to website viewers. Students have to retrieve all the information to find jobs they want to apply for. The whole procedure is tedious and inefficient. By creating an easy job recommendation system where everyone will have a fair and square chance. This saves a lot of potential time and money both on the industrial as well as the job seeker's side. Moreover, as the candidate gets a fair chance to prove his talent in the real world it is a much more efficient system. The basic agenda of every algorithm used in today's world be it a traditional algorithm or a hybrid algorithm is to provide a suitable job that the user actually seeks and wishes for

Challenges of CF-:

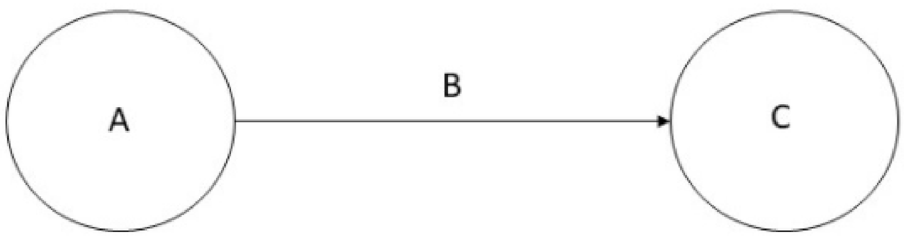
1. Scalability:-
   As the numbers of users and items grow, traditional CF algorithms will suffer serious scalability problems.Building a scalable recommendation system for millions of users and jobs is crucial. In recent years, item-based recommendation systems have gained more popularity as they are more scalable compared to their user-based counterparts. However, with the vast amount of incoming jobs everyday, building and maintaining a job-based system is not trivial. We propose a graph-based structure in order to efficiently model job-job relationships with variable-length neighborhood sizes.

2. **Job Similarities/Sparsity:-**
   In practice, many commercial recommender systems are based on large datasets. As a result, the user-item matrix used for collaborative filtering could be extremely large and sparse, which brings about challenges in the performance of the recommendation.As collaborative filtering methods recommend items based on users' past preferences, new users will need to rate a sufficient number of items to enable the system to capture their preferences accurately and thus provides reliable recommendations.Similarly, new items also have the same problem. When new items are added to the system, they need to be rated by a substantial number of users before they could be recommended to users who have similar tastes to the ones who rated them. The new item problem does not affect content-based recommendations, because the recommendation of an item is based on its discrete set of descriptive qualities rather than its ratings.

3. **Cold-Start:-**
   Relying on behavioral data solely results in lower quality recommendations when there is no user- job interaction data. This is akin to the ’cold-start’ problem which becomes more prominent in dynamic systems such as job recommendations where new users and/or jobs are introduced to the system at high rates. Moreover, many users become inactive when they do not interact with the jobs for a considerable period of time. To get the attention of such users and get them to re-engage with the system and become active users again, it is necessary to distinguish between them and brand new users.

Knowledge Graph Recommender System:-

Knowledge graph is a representation of data with the help of nodes and edges, where each of them have well defined meaning. Knowledge graphs have also started to play a central role in machine learning as a method to incorporate world knowledge, as a target knowledge representation for extracted knowledge, and for explaining what is learned.
It is far too easy to bury knowledge in documents and in heaps of natural language, and very hard to surface it at the right time. A knowledge graph makes facts easier to index, process, and find.
Defining relations and collaborating data from multiple sources is the biggest advantage of knowledge graphs.
B edge defines a relationship between nodes A and C.

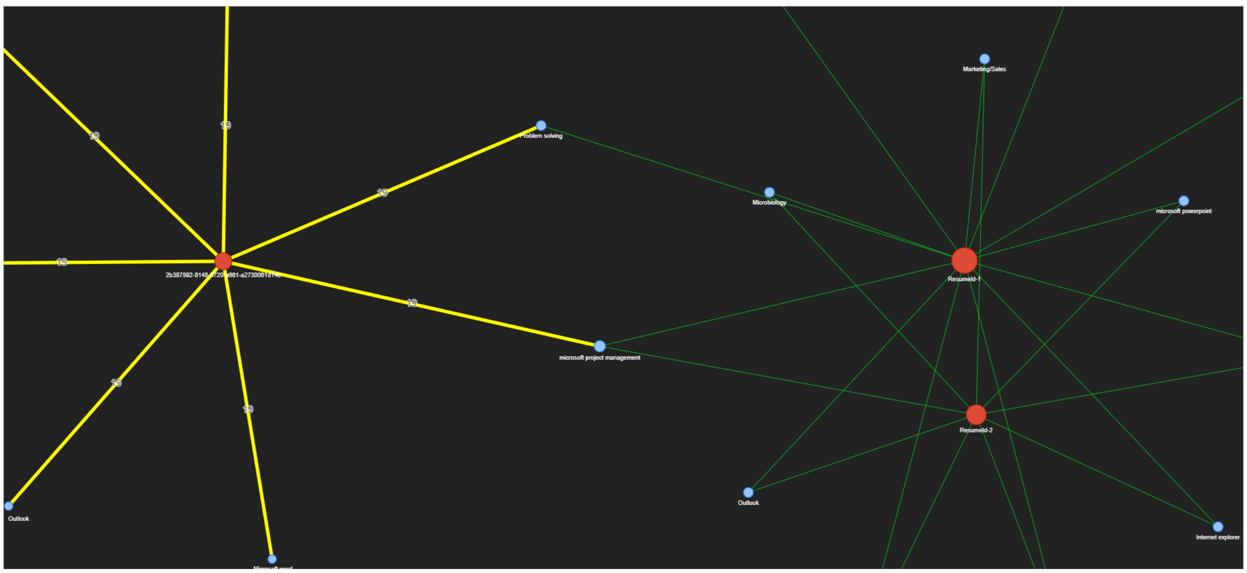


Due to this Semantic design, scalability of data is much easier. Graph-based models adopt link analysis methods from graph theory to address the shortcomings of CF-based approaches such as sparsity and improve the quality of the recommendations. As the graph is always updating with new information and forming better connections than traditional methods, we can deal with Sparsity problem to some extent.
Cold start problems are dealt in the same way, as soon as a new user node is added it will find connections with the already existing information. Rather than waiting for others to add connections with the new user.
Graph-based recommendation systems are differentiated based on how they build the graph and traverse it for recommendations. Heterogeneous graph- based models build a bipartite graph of both users and items, while homogeneous models only include users or items as nodes.

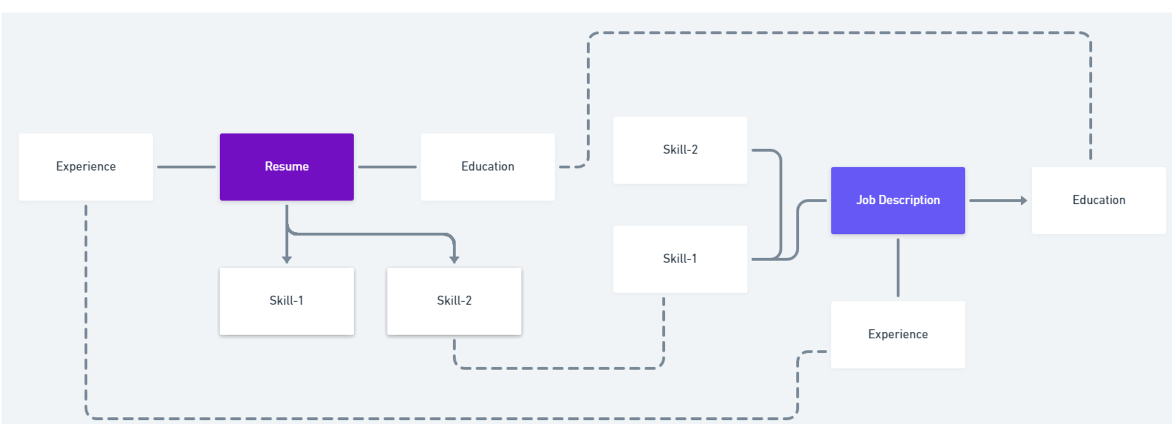
Job Recommendation System:-

Information extraction from resumes and job descriptions is one of the main areas of research in the jobs and recruitment industry. These works involve text mining, skill normalization, and developing similarity metrics for matching jobs and candidate profiles. Most of the existing recommendation systems in this domain focus on candidate selection by human resources rather than attracting job seekers through job recommendations. Existing automated job recommendation systems belong to either content-based applicant-job matcher approaches or user-based methods.

Nodes and Relationships



- The graph is bidirectional and Bipartite.
- We need 2 datasets, Job descriptions and Resume.
- We are using already existing structured data for our knowledge graph.



1. **Resume node:**- The resume node is Red in colour and has following nodes connected to it. Each resume node is assigned an ID. The green edge is used to connect the related information nodes.
    a. Skills
    b. Education
    c. Experience
2. **Job Description Node:**- The job Description Node is Red in colour and Yellow colour is used for Edge connecting related information and a particular Job Description. It has following nodes connected to it.
    a. Skills
    b. Education
    c. Experience
3. **Skills Node:-** It is Blue in colour and each node defines individual skill. It can be connected to both job Description and resume. This relation is used to derive results.
4. **Experience:**- It is Blue in colour.
5. **Education:**- It is also Blue in colour.

Prioritizing Results

1. The Resumes which have the maximum number of edges connected to a Job description. The connections are through the various Blue nodes i.e. skills, experience and education.
2. This pattern

    resume ⇔ Skills ⇔ Job description

    Means the resume is connected to Job Description and number of connected edges would count as 1.

Queries that can be answered

1. How many resumes are connected to a particular skill node?
2. Which skill is demanded most by the Job Descriptions?
3. What kind of entity is a certain term/keyword? (a skill, a degree etc)
4. What skills are required for a particular job?

And many more..!

Dataset

Job Description from dataworld

career_builder_jobs_10501.json

Query

	salary	domain	education	crawled_at
1	$17.23 - $22.00 / hour	https://www.careerbuilder.com/	Bachelor's Degree	2021-05-05T03:01:49
2	$19.00 - $26.00 / hour	https://www.careerbuilder.com/	Graduate Degree	2021-05-05T03:01:50
3	$106,250.00 - $125,000.00 / year	https://www.careerbuilder.com/	Bachelor's Degree	2021-05-05T03:01:50
4	No data.	https://www.careerbuilder.com/	High School	2021-05-05T03:01:50
5	$85,000.00 - $120,000.00 / year	https://www.careerbuilder.com/	4 Year Degree	2021-05-05T03:01:50
6	No data.	https://www.careerbuilder.com/	High School	2021-05-05T03:01:50
7	$53,041.00 - $120,750.00 / year	https://www.careerbuilder.com/	High School	2021-05-05T03:01:50
8	No data.	https://www.careerbuilder.com/	4 Year Degree	2021-05-05T03:01:50
9	$45,000.00 - $99,000.00 / year	https://www.careerbuilder.com/	No data.	2021-05-05T03:01:51
10	No data.	https://www.careerbuilder.com/	High School	2021-05-05T03:01:51
11	$65,000.00 - $120,000.00 / year	https://www.careerbuilder.com/	High School	2021-05-05T03:01:51
12	$60,000.00 - $200,000.00 / year	https://www.careerbuilder.com/	No data.	2021-05-05T03:01:51
13	$45,600.00 - $53,000.00 / year	https://www.careerbuilder.com/	High School	2021-05-05T03:01:51
14	$85,000.00 - $146,100.00 / year	https://www.careerbuilder.com/	Other	2021-05-05T03:01:52

Resume dataset link


```
[
  {
    "id": 1,
    "careerjunction_za_historical_jobtitles": [
      "Marketer & Technical Liaison",
      "Quality Assurance Manager Haccp Team Leader",
      "New Product Developer Technologist",
      "Food Technologist",
      "Quality Controller"
    ],
    "careerjunction_za_primary_jobtitle": "Senior Food Technologist",
    "careerjunction_za_employer_names": ["Cape Herb & Spice", "Greys Marine", "Heinz Foods", "Swift Silliker", "Zemcor"],
    "careerjunction_za_skills": [
      "Microbiology",
      "microsoft powerpoint",
      "microsoft office",
      "microsoft excel",
      "microsoft project management",
      "Microsoft word",
      "Outlook",
      "Internet explorer",
      "Marketing/Sales",
      "Quality Control",
      "Quality Assurance",
      "Research and development",
      "Problem solving"
    ],
    "careerjunction_za_courses": ["Btech: Food Technology", "National Diploma: Food Technology", "Senior Certificate"],
    "careerjunction_za_recent_jobtitles": ["Food Technologist", "Product Specialist Microbiology"],
    "careerjunction_za_future_jobtitles": ["Food technologist", "New product development", "auditor", "inspections"]
  },
```

Challenges

1. Dataset: To find structured dataset of resume and job description
2. Identification on Nodes: To figure out entities and nodes, weights
3. Prioritizing: Prioritize results among similar recommendations
4. Visualization: Mapping the data in knowledge graph
5. Querying: Query desired job or candidate from knowledge graph
6. Metrics for test: Check feasibility of the model

Research paper details:

- https://arxiv.org/abs/1801.00377 - Accepted at 2017 IEEE International Conference on Big Data
- http://ceur-ws.org/Vol-2947/paper17.pdf - Discussion Paper about GUApp: Enhancing Job Recommendations with Knowledge Graphs
- https://medium.com/mlearning-ai/building-a-knowledge-graph-for-job-search-using-bert-transformer-8677c8b3a2e7 - WRITTEN BY Walid Amamou
- https://engineering.linkedin.com/blog/2019/04/ai-behind-linkedin-recruiter-search-and-recommendation-systems - The AI Behind LinkedIn Recruiter search and recommendation systems
- https://www.ideals.illinois.edu/bitstream/handle/2142/98865/2pt12_Li-Job.pdf?sequence=1&isAllowed=y - Enter a Job, Get Course Recommendations accepted at iconference 2017