JRHDLSI: An Approach Towards Job Recommendation Hybridizing Deep Learning and Semantic Intelligence

Gerard Deepak[1,*], Shivam Sawarn[2] and Sheeba Priyadarshini[3]

[1]*Department of Computer Science and Engineering, Manipal Institute of Technology Bengaluru, Manipal Academy of Higher Education, Manipal, India*
[2]*School of Computing, DIT University, Dehradun*
[3]*Department of Data Sciences, CHRIST (Deemed to be University), Bengaluru, India*

Abstract

The requirement of the job for people and employees for employers are al-ways in demand. This is due to the lack of proper infrastructure to reduce the unmatching job application for employers and inappropriate job recommendations for people. This chapter proposes a strategic framework with machine learning and knowledge integration to increase accuracy in the provided recommendations and increase the chance of getting a job offer. The usage of 'user's search data intends job recommended more in liking of the users, and the machine learning helps in finding the accurate job recommendation. The machine learning technique used here is Radial Basis Function Neural Net-work for the classification and Knowledge Integrated using Analysis of Variance - Web Point Wise Mutual Information and Kullback Leibler (KL) divergence. All the job providers ads are retrieved from the top websites using beautiful soup. The proposed JRHDLSI architecture achieved an accuracy of 94.99% which outperformed the baseline models and was much superior.

Keywords

ANN, Ontology, Semantic Web, Video Classification.

1. Introduction

The Rapid growth in population in major cities has led to growth in unemployment. This is partially due to the lack of infrastructure for an efficient job recommendation system. The one that recommends jobs and does its tailor to a user is a win-win for both the employee and the employer. With more appropriate applicants and less unmatched applications, the employees can quickly speed up the hiring process and continue hiring for more roles. Machine learning techniques can efficiently handle this problem and with valid user data like the user intends so that the recommended job is appropriate and relevant. With the help of Knowledge Integration, we combine both Analysis of Variance (ANOVA) – Web Pointwise Mutual Information (PMI) and Kullback-Leibler(KL) Divergence to get an efficient model. There

DLQ-2022: International Workshop on Deep Learning for Question Answering, Co-located with the KGSWC-2022, November 21–23, 2022, Madrid, Spain
*Corresponding author.
†The authors contributed equally.
gerard.deepak.christuni@gmail.com (G. Deepak); shivamsawarn15@gmail.com (S. Sawarn); pdj.sheeba@gmail.com (S. Priyadarshini)
0000-0003-0466-2143 (G. Deepak)

are various job recommender systems and apps for a user; one does excel other in a particular category like relevance, success rate, etc. However, all users should be given equal opportunity, not the person using their time to go through 10 different job websites to get the best one. Instead, all web-site data is combined with user data and intended along machine learning to pro-duce the best results. Machine learning can give results equal to combining the best feature of different job recommending websites and apps with a fraction of time and less effort. Along with the help of Machine learning, Knowledge Integration and Accurate user data and intend the result would be tailored to the user will help reduce the overall unemployment rate.

Contribution: This chapter proposes a strategic technique for recommending jobs to users using machine learning and Knowledge Integration of ANOVA – Web PMI and KL Divergence. The Internet contains an immense amount of data regarding jobs and details of users looking for a job with a user profile that already got the job. The data are scraped using beautiful soup and preprocessed using Natural Language Processing (NLP) modules to make it machine-understandable. With these data and Radial Basis Function Neural Network (RBFNN), we classify data under different categories. These are then Knowledge Integrated with user profile and search intends using ANOVA -Web PMI and KL Divergence.

Organization: The next part of the chapter is as follows: The second section addresses the relevant research previously done related to this topic. The third section explains the architecture for the proposed system in brief. Section four consists of the proposed model's implementation and the performance evaluation. Section five presents the conclusion of the chapter.

2. Related Works

Ephizibah et al. [1] have put forth a framework that recommends jobs using a deep learning model, which uses standard user data, job requirement details, geographic location, and employee database. Based on existing studies, siting et al. [2] have explored basic concepts of standard job recommendation technologies and user profiles. Diaby et al. [3] have proposed a job recommendation frame-work for Facebook and LinkedIn users based on the content. Hong et al. [4] has proposed a job recommendation system that uses the employee's job description and job search query of the user to recommend job using clustering.

Hong et al. [5] have suggested an adaptive framework for user profile-based job recommender in which all the user profile is dynamically generated based on previously applied jobs and job applicant behaviours. Lee et al. [6] have put forth a comprehensive job recommender framework that implements four different types of recommendations. Gupta et al. [7] have put forth a system for candidate information and behavior to recommend jobs. Liu et al. [8] have put forth a system that combines temporal learning and sequence modelling to capture complex user-item interaction patterns and refine job recommendations.

Mpela et al. [9] have proposed a mobile job employment recommender frame-work. This client-server software employs a content-primarily based filtering set of rules to permit the initial choice of an appropriate enjoyment task seeker for a transient task at a selected vicinity and vice versa. Tayade et al. [10] have looked at some standard work recommendation systems and a data mining approach to job recommendation systems.

Zhang et al. [11] have put forth a collaborative filtering algorithm based on user and item to evaluate which is the better per-forming one. Yang et al. [12] have put forward a model to use SRL (Statistical Relation) to combine content-based filtering and collaborative filtering, producing a hybrid job recommender system. Ortiz-Rodriguez et al., [13] have formulated the MEXIN which is a multi dialectal ontology which provides linguistic corpora to achieve NLP support with a focus on improvisation of electronic communication between several Mexical ethnic groups via cognizable Ontological entities. Gupta et al. [14] have proposed a novel model for facilitating question answering in an environment of a highly domain dependent knowledge graph which focuses on Indian Missiles. The principle of reasoning over knowledge graphs using Neo4j platform is achieved over a knowledge graph with varied formal entity set. This work is a confidence booster of how Semantic Reasoning can derive insights over a information dense unit like the Knowledge Graph.

Abhishek et al.[15] have put forth an intelligent model for mining knowledge graphs for online news by entity extraction over non-trivial knowledge pockets over a highly dynamic Web 3.0. The model stabilizes the knowledge graph formalization over a highly changing news environment for mining news via knowledge that has been synthesized thereby accelerating Semantic Intelligence over a Knowledge Graph unit. Phukan et al. [16] have synthesized a stress recognition for digital healthcare for facilitating e-governance using wearable sensor devices and capturing data which is transformed into features over a transformer based deep learning model to achieve machine intelligence via knowledge representation schemes. Usip et al. [17] have also put forth a personal profile Ontology to ease Software Requirements Engineering allocation of tasks. The proposed Ontology model captures both static and dynamic data properties and also mixes Ontological Strategies like Neon and Methontology along with e-PPO model for achieving dynamism over ontological properties for task allocation and reasoning. These literature gives a confidence on how knowledge representation and reasoning over the represented knowledge can achieve cognizable machine intelligence via Semantic Frameworks over highly cohesive and dynamic knowledge units.

3. Proposed System Architecture

An intelligent job recommendation service has been explained in this section, as shown in Figure. 1. The system which has been proposed consists of six phases, data crawling, metadata generation, classification using RBFNN, extraction term from user's job profile, knowledge integration, the recommendation for job seeker. Data is collected using Beautiful Soup. Beautiful soup is a Python package that allows you to analyze structured data. It allows researchers to go through the website in python similarly to the inspection option available in the browser. Beautiful soup has some built-in functions for studying HTML. Beautiful soup is used on the popular job-seeking website to retrieve all job details and descriptions.

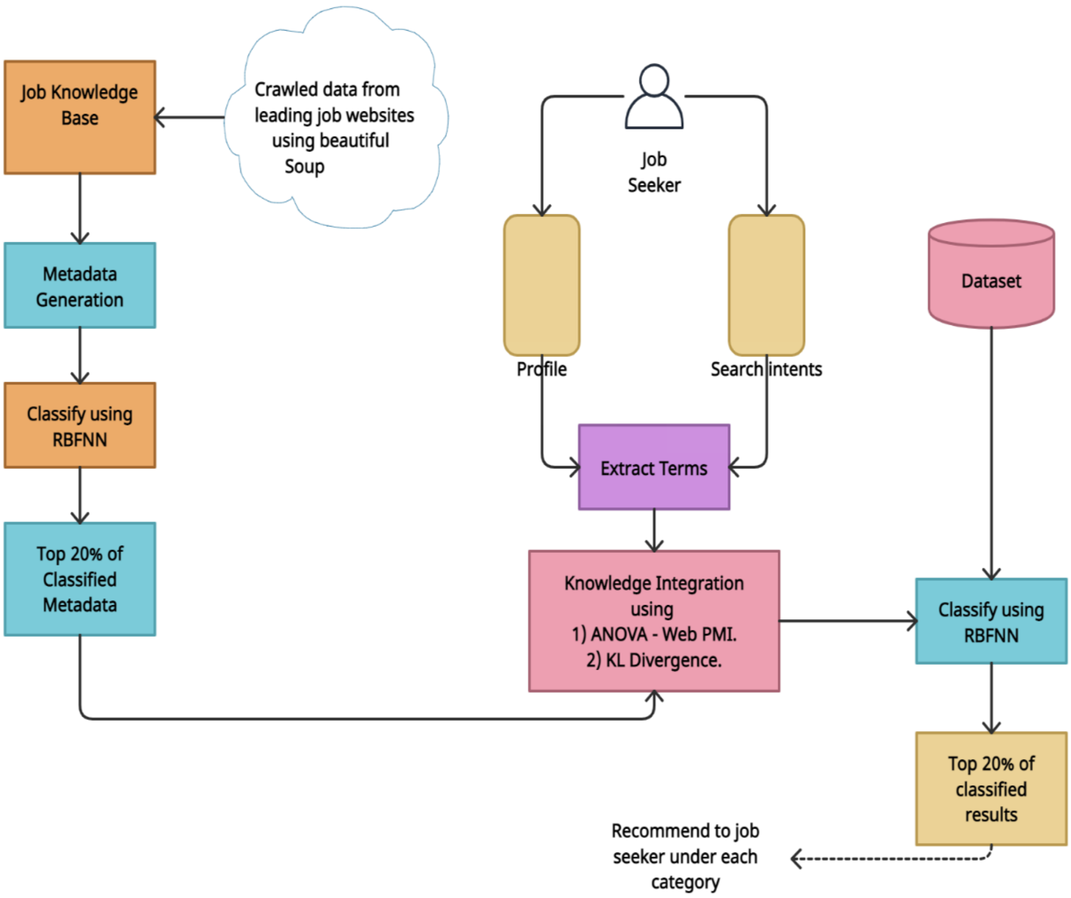

Figure 1: Proposed System Architecture

The process of extracting additional information from the Web resource concerning ontology is called metadata extraction, which is required for our experiments. We use it for the crawled data to extract metadata out of it. Which is processed using NLP tools like word tokenize, stop words removal, normalization, context-free grammar for words related to 'job', stemmed and lemmatized. These words are one-hot-encoded to fit for the model.

The RBFNN is a three-layered neural network that is feed-forward. The first layer is linear, distributing only the input signal, while the second layer is nonlinear, using Gaussian functions. The Gaussian outputs are linearly combined in the third layer. During preparation, only the tap weights between the hidden and output layers are modified. Then the whole metadata is again classified, and the accuracy orders the top 20% of each category.

The statistical method of analysis of variance (ANOVA) is broadly divisible into two parts based on the overall variability: system components and random factors. While the system components have an impact statistically on a given dataset, the random factors do not. It gauges the influence of the independent variables on the dependent variable. The t and z test were developed in the 20th century and was used for statistical analysis until Ronald Fisher introduced the analysis of variance (ANOVA) in 1918. It first found its use in experimental psychology and then expanded to include other topics which were complex. In Equation (1), F is the ANOVA coefficient, MST is the sum of mean squares caused by processing, and MSE is the sum of mean squares caused by errors.

$$F = \frac{MST}{MSE} \tag{1}$$

The Kullback-Leibler divergence or the KL divergence determines the degree of difference between one probability distribution and another. The KL divergence is usually given by two distributions Q and P, represented by Equation (2). KL divergence is evaluated by the negative total of the probability of every event present in P increased by the index of the probability of the event present in Q and the probability of events in P. The probability present in P is multiplied by the logarithm of the probability of the event in Q and the probability of events in P.

The intuition of KL divergence is that once the probability of a specific event at P is high and a particular event at Q is low, there will be an enormous divergence. If the probability P is of a smaller value, then the probability Q is more extensive. There exists a vast difference. However, not as significant because of the initial case. It will be accustomed to the widths of discontinuous and continuous probability distributions to calculate the case integral instead of the total of the possibilities of fragmented, separate events.

$$D_{KL}(P||Q) = \sum_{x \varepsilon X} P(x) log \left(\frac{P(x)}{Q(x)} \right) \tag{2}$$

Which is processed using NLP tools like word tokenize, stop words removal, normalization, context-free grammar for words related to 'job', stemmed and lemmatized. These words are one-hot-encoded to fit for the model. The artificial recurrent neural network (RNN) architecture is used in the long short-term memory (LSTM) machine learning framework. LSTMs have feedback linkages, unlike standard feed-forward neural networks. It can handle complete data sequences as well as single data items (speech, etc.).

4. Implementation

The implementation of the proposed approach is done using the python language and Jupiter notebook IDE. There are 6 phases in the process of implementing the proposed system. The tools and library in the implementation of proposed architecture beautiful soup, sklearn and NLTK. It is inferable that the performance exhibited by the proposed system is computed by the usage of precision, F-measure, recall, precision, accuracy and False Discovery Rate (FDR) as the potential metrics. The recall is the proportion of ontologies recovered and applicable to the total number of relevant ontologies. Precision is characterized as the proportion of the retrieved and significant ontologies to the overall number of recovered ontologies.

For precision and recall measures, accuracy is specified as the average. The Accuracy encompassed in this case is the Average Balanced Accuracy which is the mean of Precision and Recall percentages. The False Discovery Rate (FDR) quantifies the number of False Positives furnished by the framework. Equations (3), (4), (5), (6) and (7) depicts the Precision, Recall, Accuracy, F-Measure and the FDR. In order to evaluate, quantify and compare the performance of the proposed JRHDLSI framework, it is baselined with Clustering with Deep Learning [1] , Content-Based Filtering+ Collaborative Filtering [11], and Collaborative Filtering [12],

$$Precision = \frac{No.\ of\ recovered\ and\ pertinent\ Ontologies}{Total\ No.\ of\ Ontologies\ Recovered} \tag{3}$$

$$Recall = \frac{No.\ of\ recovered\ and\ pertinent\ Ontologies}{Total\ No.\ of\ Ontologies\ Significant} \tag{4}$$

$$Accuracy = \frac{Precision + Recall}{2} \tag{5}$$

$$F - Measure = \frac{2 * P * R}{(P + R)} \tag{6}$$

$$FDR = 1 - \frac{Precision\%}{100} \tag{7}$$

From Table 1 it is inferable that the Proposed JRHDLSI framework yields the highest average percentages of Precision, Recall, Accuracy, F-Measure and the lowest value of FDR. The main reason for the proposed JRHDLSI to outperform the baseline models is primarily due to the reason that it is a semantically inclined model which is powered by anchoring auxiliary knowledge. The auxiliary knowledge is selectively encompassed by means of automatic generation of Metadata and subject to its classification by RBFNN. The reason for classification of the metadata is due to its exponentially large scale and knowledge be harvested and regulated from metadata, which is seasoned using ANOVA-Web PMI Model and the K-L Divergence measure. The ANOVA-Web PMI sandwich ensures the computation of the pointwise mutual information measure for the entities in the Web Corpus and the ANOVA model ensures an implicit thresholding scheme for filtering out irrelevant entities and instances and retaining the entities and instances that are relevant. Also, the KL Divergence measure set to an implicit step deviance criterion ensures further filtering out entities between the classified metadata and the entities in the user intents and user profile. This further helps in ensuring the most relevant entities and instances be retained into the final recommendation list increasing the strength of relevance computation. Moreover, the RBFNN is also employed to classify the Dataset just as the Metadata and this is done as a regulatory scheme for filtering out instances

Table 1
Comparison of Performance of the proposed JRHDLSI with baseline model approaches

Search Technique	Average Precision %	Average Recall %	Accuracy %	F-Measure %	FDR
Clustering with Deep Learning [1]	85.41	88.25	86.83	86.81	0.15
Content-Based Filtering+ Collaborative Filtering [11]	84.63	86.12	85.37	85.37	0.16
Collaborative Filtering [12]	81.17	83.44	82.30	82.29	0.19
Proposed JRHDLSI (Metadata+ RBFNN+ANOVA Web PMI with KL Divergence)	93.28	96.71	94.99	94.96	0.07

and assure computationally inexpensive and highly relevant final recommendations. Moreover, the RBFNN is a strong learning strategy and serves as a fairly strong classifier with a high learning rate for the computational expense at this scale.

The reason why Collaborative Filtering as a standalone strategy does not perform fell in yielding highly relevant recommendations is mainly due to the fact that it works on the principle of item similarity computation using a matrix formation strategy, and requires a strong rating paradigm by a community at large. It is impractical for achieving a uniform consensus and rating for every specific job and it is not a fair criterion for deciding on the relevance for recommendation. The Collaborative Filtering technique [12] yields the lowest Precision % of 81.17, the lowest Recall % of 83.44, lowest Accuracy % of 82.29 and the highest value of FDR of 0.19. However, when Collaborative Filtering was coupled with Content Based Filtering, it yielded an overall Average Precision % of 84.63, overall Average Recall % of 86.12, overall Average Accuracy % of 85.37, overall Average F-Measure 5 of 85.37 and a FDR of 0.16. However, when content based filtering was coupled with Collaborative Filtering, the relevance improved as the contents provisioned concrete substrate for the relevance computation measures to act upon. However, still the combination of content based filtering with Collaborative Filtering does not serve as a win-win model as still rating matrix computation do not go along with relevance computation strata.

The hybridization of Clustering with Deep Learning [1] paradigm does not yield the best-in-lass results and still lags by yielding an overall Precision % of 85.41, an average percentage Recall of 88.25, an average Accuracy % of 86.83 and an average percentage of F-Measure of 86.81 with an FDR of 0.15 due to the fact that although Deep Learning is a strong classification scheme, the model lacks useful relevant auxiliary knowledge which results

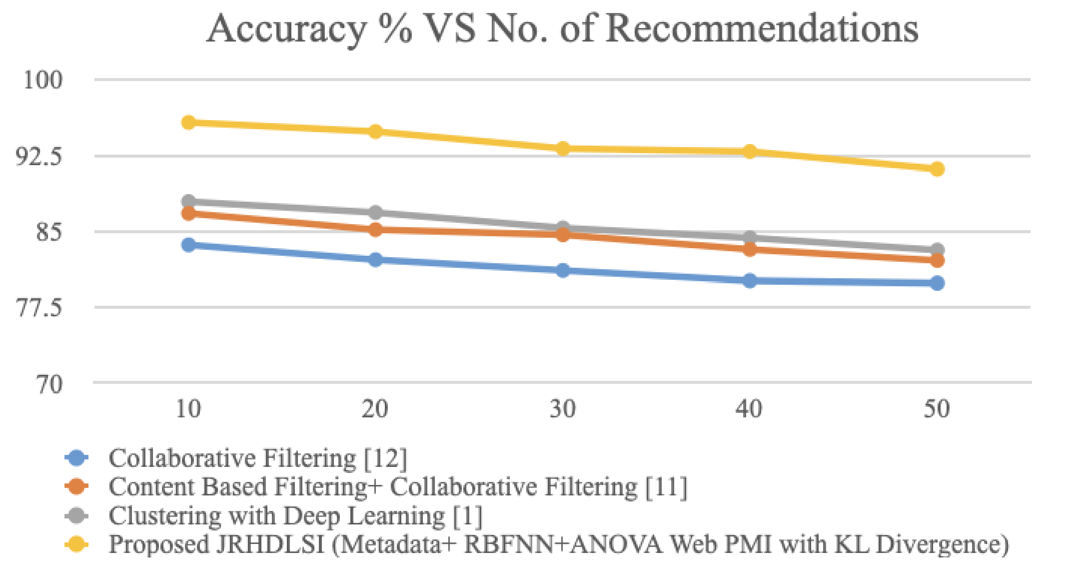

Figure 2: Accuracy Vs No. of Recommendation

in over specialized learning with under fitting of useful knowledge. Even on coupling of Clustering with the Deep Learning paradigm, this model lags due to the scarcity of auxiliary knowledge. Owing to the lacunae in the baseline models and the presence of Metadata for harvesting exponential scale of auxiliary knowledge which is seasoned using the RBFNN classifier for transforming into relatable knowledge capsule based binding into the proposed model. The integration of ANOVA Web PMI with KL Divergence measure assures strong relevance computation scheme in the environment of surplus, concrete and dense knowledge in the proposed JRHDLSI. Owing to the hybridization of colletive intelligence with auxiliary knowledge in the proposed JRHDLSI framework, it furnishes the highest Precision % of 93.28, the highest Recall % of 96.71, highest Accuracy % of 94.99, highest F-Measure % of 94.96 with the lowest value of FDR of 0.07.

Figure. 2. shows each model's performance at 10, 20, 30, 40 and 50 recommendations for Collaborative Filtering, Content-Based Filtering added with Collaborative Filtering, Clustering with Deep Learning and JRHDLSI (proposed approach). The data given across all the models are the same and given in the same order. There is a performance dip, and that is based on the complexity of the input data. Since all the datasets and orders are the same, the performance is taken to evaluate the model. Collaborative filtering has the lowest average dip, but the average accuracy is 81.39%, and the proposed approach has the second-lowest dip and has the average accuracy of 93.57%, making the proposed approach superior to just using Collaborative Filtering. The proposed model JRHDLSI has precision, recall, accuracy, F-measure and FDR is 93.28%, 96.71%, 94.99%, 94.96%, and 0.0672 respectively, has the lowest performance dip to No. of recommendation shown in Figure. 2. and has better performance metric in comparison to other three baseline model making the proposed model one of the better approaches for smart job recommendation system. The reason why the proposed JRHDLSI model occupies the uppermost position in the hierarchy is primarily due to the reason that it is a metadata centric, metadata driven strategy which classifies both the metadata and the Dataset by encompassing RBFNN which is a strong deep learning scheme for classification. Apart from this the metadata classified is seasoned as reasonable and inferable knowledge by subjecting it to the ANOVA-Web PMI model and the KL Divergence to compute the relevance of results.

5. Conclusions

The need for jobs for people and employers is always present because of the lack of efficient infrastructure connecting these two groups. An efficient model for job recommendation useful to the job seekers has been proposed so that the number of rejections for both groups are at the minimum. The strategy encompasses RBFNN for classifying metadata as well as the dataset. The proposed JRHDLSI strategy for job recommendation is a preferential knowledge driven paradigm. The strategy encompasses Knowledge integration which is obtained by classification of metadata and preferential selection using ANOVA - Web PMI and KL Divergence. This approach has a learning feature that takes the search as the input to give a more accurate and relevant recommendation dynamically. An overall F-Measure of 94.96 % has been achieved with the lowest FDR of 0.07.

References

[1] EP, E. (2019, February). Framework of an Intelligent Job Recommendation System. In Proceedings of International Conference on Sustainable Computing in Science, Technology and Management (SUSCOM), Amity University Rajasthan, Jaipur-India.

[2] Siting, Z., Wenxing, H., Ning, Z., & Fan, Y. (2012, July). Job recommender systems: a survey. In 2012 7th International Conference on Computer Science & Education (ICCSE) (pp. 920-924). IEEE.

[3] Diaby, M., Viennet, E., & Launay, T. (2013, August). Toward the next generation of recruitment tools: an online social network-based job recommender system. In 2013 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM 2013) (pp. 821-828). IEEE.

[4] Hong, W., Zheng, S., Wang, H., & Shi, J. (2013). A job recommender system based on user clustering. JCP, 8(8), 1960-1967.

[5] Hong, W., Zheng, S., & Wang, H. (2013, April). Dynamic user profile-based job recommender system. In 2013 8th International Conference on Computer Science & Education (pp. 1499-1503). IEEE.

[6] Lee, D. H., & Brusilovsky, P. (2007, June). Fighting information overflow with personalized comprehensive information access: A proactive job recommender. In Third International Conference on Autonomic and Autonomous Systems (ICAS'07) (pp. 21-21). IEEE.

[7] Gupta, A., & Garg, D. (2014, September). Applying data mining techniques in job recommender system for considering candidate-job preferences. In 2014 International Conference

on Advances in Computing, Communications and Informatics (ICACCI) (pp. 1458-1465). IEEE.
[8] Liu, K., Shi, X., Kumar, A., Zhu, L., & Natarajan, P. (2016). Temporal learning and sequence modeling for a job recommender system. In Proceedings of the Recommender Systems Challenge (pp. 1-4).
[9] Mpela, M. D., & Zuva, T. (2020, August). A Mobile Proximity Job Employment Recommender System. In 2020 International Conference on Artificial Intelligence, Big Data, Computing and Data Communication Systems (icABCD) (pp. 1-6). IEEE.Diaby, M., Viennet, E., & Launay, T. (2014). Exploration of methodologies to improve job recommender systems on social networks. Social Network Analysis and Mining, 4(1), 227.
[10] Tayade, T., Akarte, R., Sorte, G., Tayade, R., & Khodke, P. (2020, January). Data Mining Approach to Job Recommendation Systems. In International Conference on Mobile Computing and Sustainable Informatics (pp. 503-509). Springer, Cham.
[11] Y. Zhang, C. Yang and Z. Niu, "A Research of Job Recommendation System Based on Collaborative Filtering," 2014 Seventh International Symposium on Computational Intelligence and Design, 2014, pp. 533-538, DOI: 10.1109/ISCID.2014.228.
[12] Yang, S., Korayem, M., AlJadda, K., Grainger, T., & Natarajan, S. (2017). Combining content-based and collaborative filtering for job recommendation system: A cost-sensitive Statistical Relational Learning approach. Knowledge-Based Systems, 136, 37-45.
[13] Ortiz-Rodriguez, F., Tiwari, S., Panchal, R., Medina-Quintero, J. M., & Barrera, R. (2022, June). MEXIN: multidialectal ontology supporting NLP approach to improve government electronic communication with the Mexican Ethnic Groups. In DG. O 2022: The 23rd Annual International Conference on Digital Government Research (pp. 461-463).
[14] Gupta, Shivansh, Sanju Tiwari, Fernando Ortiz-Rodriguez, and Ronak Panchal. "KG4ASTRA: question answering over Indian missiles knowledge graph." Soft Computing 25, no. 22 (2021): 13841-13855.
[15] Abhishek, Kumar, Vaibhav Pratihar, Shishir Kumar Shandilya, Sanju Tiwari, Vinay Kumar Ranjan, and Sudhakar Tripathi. "An intelligent approach for mining knowledge graphs of online news." International Journal of Computers and Applications 44, no. 9 (2022): 838-846.
[16] Phukan, O.C., Singh, G., Tiwari, S., Butt, S. (2022). An Automated Stress Recognition for Digital Healthcare: Towards E-Governance. In: Ortiz-Rodríguez, F., Tiwari, S., Sicilia, MA., Nikiforova, A. (eds) Electronic Governance with Emerging Technologies. EGETC 2022. Communications in Computer and Information Science, vol 1666. Springer, Cham.
[17] Usip, P.U., Udo, E.N., Umoeka, I.J. (2021). An Enhanced Personal Profile Ontology for Software Requirements Engineering Tasks Allocation. In: Villazón-Terrazas, B., Ortiz-Rodríguez, F., Tiwari, S., Goyal, A., Jabbar, M. (eds) Knowledge Graphs and Semantic Web. KGSWC 2021. Communications in Computer and Information Science, vol 1459. Springer,