Task Recommender System using Semantic Clustering to Identify the Right Personnel

Prafulla Bafna, Symbiosis International University (SIU),

India, Maharashtra, Pune

S.C Shirwaikar, savitribai Phule Pune Univesity

India, Maharashtra, Pune

Dhanya Pramod, Symbiosis International University (SIU),

India, Maharashtra, Pune

**Abstract**

**Productivity of any organization is enhanced by assigning task to employees with right set of skills. An efficient recruitment and task allocation can not only improve productivity but also cater to satisfying employee aspiration and identifying training requirements. An automated Task Recommendation System is proposed comprising of synset based feature extraction, iterative semantic clustering and mapping based on semantic similarity. The system takes documents like appraisal or resume as input and suggests not only the persons appropriate to complete a task and job position but also employees needing additional training.**

**Keywords: Synset group, clustering, Recommender system, entropy**

**INTRODUCTION**

The success of any organization depends on the competence and superiority of its employees. It becomes essential to choose correct people and place them at right place. This requires that both the recruitment process and the task allocation process being streamlined and optimized.

Recruitment process is generally conducted by HR department of company to find out and attract capable applicants. The process provides the pool of potentially skilled candidates for the job.

It starts with determining the present and future requirements of the organization in conjunction with its personnel planning and job analysis activities. The HR team needs to identify and prepare potential job applicants who will be appropriate candidates. (Derous, 2016) The systematically conducted process increases organization increases the effectiveness for individuals in the organizations both in the the short term and long term basis.

Task allocation is an internal activity carried out in an organization in which workloads associated to the task are distributed among the available employees. Generally, task allocation is based on suggestion of senior employees who have experienced a successful task completion from their team members. Sometimes employee also suggests their aspiration or the task at which they are more comfortable. The accomplishments and aspirations are usually spelled out in self appraisal forms and can be focused while allocating the task.

HR department utilizes resumes of aspiring candidates and self appraisal forms of existing employees in their decision making process. Both being unstructured documents requires effort to extract the right information ,so text mining techniques can be helpful such as clustering, feature extraction etc. In text mining major terms (frequently occurring) are considered to represent document. (Bafna, 2016) Clustering helps in grouping documents based on the similarity between terms present in the document. Semantic similarity can be emphasized further by making groups of synonyms, meronyms etc. This paper presents holistic process that combines recruitment and task allocation process that is assisted by mining initiatives for improving the effectiveness of decision process.

A Synset based task Recommender system is proposed, which maps skillset of an employee that is the expertise pertaining to the employees to the skillset required to perform the task efficiently. The employees which are not able to fulfill the requirement of that task because of absence of skillset are recommended training so that they will be eligible to perform the next similar project task.

**BACKGROUND**

The productivity of the employees heavily depends on the effective utilization of their inherent skills and competencies and focused efforts to impart training in skills and technologies required for the growth of the organization. HR department acquires, develops, utilizes and maintains employees. Acquiring right man for the right job at right time in right quantity, developing right kind of training, utilizing the selected workforce, and maintaining the workforce, are the organizational objectives of HRM. Thus choosing the right set of employees for the given task, is the first step in this direction.( Horne, 2016)

**Screening Process**

Many screening methods are used for recruitment like aptitude round, group discussion, Personal interview etc. but initial screening is usually based on resumes.(Nunley, 2016) Resume is a valuable document of a candidate, which provides him the platform to summarize experience, training, skills and knowledge. It lists a job applicant's education and activities. It is used to provide a detailed summary of an applicant's qualifications for a particular job. A good resume is meant to provide a complete picture about the potentials of candidate and reflects whether the applicant is worth interviewing.(Derous, 2017)

Recruiters need resumes mapping the skillsets required by the vacancy in the particular department. (Jannach, 2010).

While resume is a key and essential part in securing an interview by an external candidate, at internal level, self-appraisal forms are used to describe achievements and the limitations of an employee and both together guide the task allocation process.(Risavy, 2017 )

**Recommender System**

The efforts required for abstracting or finding out important data from these documents is overwhelming in which case the candidate selection can simply be guided by recommendations by the experts. These recommendations can be from people who have actually observed the skillset of a candidate in the capacity of a supervisor or manager. Recommender systems are software systems that automate this recommendation process.

Recommender systems are widely used in on-line sales which have the ability to suggest most relevant and accurate item to the user by presenting useful stuff. It filters out the necessary information from pool of sales and user data. It predicts whether a particular user will prefer an item or not. User's profile, record of past purchases, searches and other users' behavior act as input in this prediction. Amazon uses recommendation algorithms to personalize the online store for each customer. The store updates based on customer interests. Amazon or similar on-line vendors suggest products based on previous purchases of items or some strategy. (Greg Linden et al,2007, Smith, B., & Linden, G. , 2017). Netflix gives recommendations of movies based on ratings provided by users. The first commercial recommender system Tapestry, (Gomez-Uribe,2016) was designed to recommend newsgroup documents. Recommending can help users to focus on the relevant news instead of getting confused by the bombardment of too many news documents. (Miller, 2003) (Melville, 2009)

The two main techniques for RS are Collaborative Filtering systems that analyze historical data of wider set of Users and Content-based Filtering systems that are based on profile attributes of the User and the attributes of the item to be recommended. Hybrid (Jashki,2009) techniques combine both of these designs.(Debnath,2008) (Hung, 2012)(Wei, T.,2015)

In Content filtering, user profile data is used to recommend items. Attributes present in the user profile are mapped. The user’s choice and preferences are gathered and stored. The features of the content item are also extracted. Recommendation of the specific item is done to particular user by matching the features of items to user’s preferences.

Collaborative filtering predicts the degree to which customer will prefer the product. It is based on attributes of item similar to other items chosen by similar customers.( Zhang, F., Yuan, N. J., Lian, D., Xie, X., & Ma, W. Y. , 2016). To find out similar products and users, clustering is used. It facilitates to present few top recommendations. (Glance,2001;Karypis,2001;Kumar,2014)

RS provide personalized shopping by presenting items that were selected by similar customers. RS provides the personal attention usually found in the physical stores.

Content filtering is based on favorite items where as preferences of users to products is used by collaborative filtering. Matrix factorization is widely used approach when product information in the form of user rating is available.(Luo, X., et.al, 2016) (Li, H et.al, 2017) .CLEF NewsReel uses deep leaning to recommend large scale online news . (Kille, B., Lommatzsch, A., Hopfgartner, F., Larson, M., & Brodt, T. , 2017).

**Semantic Relativity Impact**

The recommender system should be able to search suitable candidates for the required job positions. Resumes pertaining to different domains contain the specific keywords e.g the resume for the accountant post may have specific keywords like Tax audit, recovery etc, These keywords are used to retrieve the resumes required by the recruiters. Recruiters require selecting the person matching candidate’s experience with business needs. Resumes in the form of textual data is unstructured and need to be converted into vector form known as term document matrix in which rows represent resumes and columns represent important features or terms extracted from resumes. It does not use semantic relevancy of words and in absence of such feature, similar words get placed in different columns. (Jayabharathy, 2014)

Resumes describe the personality of a person, job profile and expertise. It also represents the summary of talents, capabilities and endeavors. It gives the summarized view of a person. Recruiters are interested to record the snapshot of a person with the intent of selecting candidate for interview. Same Resume written by two humans may contain different words for the same concept. The words describing the same contexts are generally semantically related. Semantic similarity measure can be effectively applied by grouping the terms into synsets as provided by WordNet. WordNet is the most extensive and widely used thesaurus providing groups of noun, verbs, adjective and adverbs called synsets. Synsets are organized using semantic relationships (Pantel, 2002)(Wei,2015) such as synonyms, hyponyms, hypernyms, meronyms and holonyms. Thus column in the term document matrix (Li, 2009) are replaced by synsets which tends to improve the selection of major terms.

**Evaluation of clusters**

Clustering is used for grouping similar documents based on proximity of extracted features or terms. Generally cluster quality is measured using F-measure, entropy and precision. Entropy measures the uniformity or purity of a cluster and precision directly reflects the performance of clustering. Entropy is used with various preprocessing methods such as wrapper, filter for feature elimination, reduction and selection. (Jashki,2009 ; Rafi, 2013)

A well-balanced coefficient, the silhouette width, is a good measure to validate the number of clusters. It was introduced by Kaufman and Rousseeuw. It is the difference between the within-cluster tightness and inter-cluster separation. The silhouette width values are from 0 to 1. Higher average silhouette width indicates better clustering. (Mark Ming-Tso Chiang, 2009 ; Mark Ming, 2009 ;Herlocker, 2004; Halkidi,2002; Faridani, 2011)

**TASK RECOMMENDER SYSTEM**

Recommendation process is used to recommend persons based on the skillset required to complete a task or to fulfill the job position. In a small set up, a senior manager may recommend juniors based on his knowledge about the persons working under him. This many times is judgmental or biased and does not cater for the aspirations of employee. The skills and experience information about the person is present in resumes or appraisal forms which are unstructured documents. An automated Recommendation process can extract the best set of people matching the skillset and also spell out the unmapped or leftover skillsets without appropriate matches as well as the people left behind without matching tasks. The Task recommender system as an input , process, output system is shown in figure1.

**Unstructured data in the form of documents**

**Required skillset**

**Task Recommendation System**

**Unmapped people**

**Unmapped skillset**

**Recommended people**

Figure 1 : Block diagram of Task Recommender System.

In any organization responsibilities and workloads related to one task are distributed among different individuals according to expertise of those individuals. The Skillset expected to complete that task can be extracted. Appraisal forms contains the task willingness of employee towards the work, expertise gained from earlier assignments and acquired skillsets. The task recommender system can be used at this stage to get the recommendations as also unmatched skillset and employees. The recommendations can be input to the decision process that assigns task to employees. The unmapped employees should be given training in unmapped skillsets or new employees should be recruited. For recruitment process task recommender system can be used again to match the unmapped skillset to resumes of newly entering candidates. Figure 2 shows the complete flow of task allocation in an organization using task recommender system

**Initial Feature extraction**

**Hierarchical clustering**

**Cluster Validity**

**Extracting feature set at the lowest level of clusters**

**Final Extracted Features**

**Mapping**

**Unmapped Skillset**

**Unmapped**

**People**

**Recommended**

**People**

**Documents**

**Skillset**

**Extends**

.

Figure 2: Task allocation in task recommender system

The importance of the Task recommender system is evident from the above workflow as it is used both for internal task allocation as well as selection of new employees. A task recommender system based on semantic similarity is described below.

It comprises of three basic steps known as feature extraction, iterative clustering, and similarity mapping. The flow of three step approach is shown in figure 3.

**Recommendation Process**

**Recommended Resumes**

**Unmapped skillset**

**Get candidate Resumes**

**Unmapped**

**Employees**

**Decision**

**Process**

**Recommendation**

**Process**

**Training**

**Identify task skill set**

**Get appraisal forms of employees**

Figure 3: Flow of three step approach in task Recommender System.

Task recommender system has two basic Components feature extraction and mapping. Feature extraction can be iterative or non-iterative.

Skillset required for a job position can be specified through a single document or a set of documents ,for example for a job of accountant a skillset can be provided as a single document or set of best resumes of good accountants which together indicate required skillset.

Feature extraction can be applied to set of documents which will group the document and extract the features of clusters representing specific skillset

The Term Document matrix contains the frequency of occurrence of the term. The synset grouping of terms is based on both synonyms and meronyms. Dimension reduction is applied by selecting the major terms that is the terms with frequency higher than the threshold. The steps are described below and the dendrogram of final clusters obtained on a smaller size sample (24 records) is shown in Figure 4



Figure 4 : Dendrogram for 24 resumes.

**SYSTEM IMPLEMENTATION AND VALIDATION**

The efficacy of the proposed approach is validated by performing several experiments. The experiments are conducted using different packages available in R programming e.g. TM, Snowball, StringR etc. R-Wordnet is used for incorporating semantic relativity between the terms.

Available data on the web is used to perform large scale experiments. Various websites provides the resume like career cloud, freeresumesites etc. The dataset used was the Corpa of resume belonging to the different domains like accounts, sales, purchase, customer care etc. Resumes were preprocessed to get the Term document matrix

**Iterative Feature Extraction**

If a skillset is specified by single doc directly features can be extracted in the form of majority synset groups while set of documents are provided for different skillset then iterative semantic clustering can be used . Different steps to carry out Iterative feature extraction are listed below.

1.Getting repository of documents.

2. Pre-process the documents/ skillset (stemming, stop words removal, correct misspelled words)

3. Extract terms from documents/ skillset(feature vector)

4. Identify synonym for each term in the term set of documents and group them.

5. Get the group frequency of each synset group.

6. Consider the synset groups having higher or equal count than threshold

7. Formulate the term document matrix

The feature set size increases as the larger set of resumes are processed. Table 1 shows the feature set dimensions on different size samples. The features can be extracted for each cluster. Each feature set is a synset group and is represented by its parent term.

Table 1: Feature set dimensions

|  |  |  |  |
| --- | --- | --- | --- |
| Number of Resumes | Dimensions | Number of clusters | Entropy |
| 24 | 20 | 3 | 0.063 |
| 49 | 35 | 4 | 0.066 |
| 200 | 40 | 4 | 0.061 |
| 600 | 42 | 4 | 0.092 |
| 800 | 42 | 4 | 0.091 |
| 1000 | 45 | 4 | 0.096 |

**Iterative Refinement Process**

Clustering is a popular descriptive data mining technique used for grouping similar objects. The grouping becomes more realistic when semantic relativity is brought into picture.

There are several clustering approaches; we have extended the most efficient hierarchical agglomerative approach by incorporating the semantic support. Hierarchical clustering approach provides several levels of clusters. The outcome can be refined by performing iterations and entropy is used to decide the number of iterations.

The hierarchical grouping is based on cosine distance and the output is in the form of a dendrogram. In the next iteration, the features are extracted at the lowermost level of hierarchy and are added to the feature set at the highest level of hierarchy. The process is repeated on this extended feature set. The cluster entropy is used to decide the number of iterations. The entropy becomes stable after finite number of iterations. The changes in the entropy, with every iteration and convergence of algorithm for 600 resumes are given in Table 2. Silhouette’s coefficient acts as internal cluster validity parameter.

Table 2: Convergence Of algorithm for 600 resumes

|  |  |
| --- | --- |
| Iteration Number | Entropy |
| 1 | 0.391 |
| 2 | 0.102 |
| 3 | 0.092 |

**Hierarchical clustering with synset grouping**

Instead of rigidly defining the k value, hierarchical clustering forms a level of clusters, with number of clusters varying at every level. Hierarchical algorithm as defined in section 3 is applied on a sample of 49 resumes, selected from 4 different domains. Resumes and retrieved cluster features are based on sales, purchase and accounts domain. Entropy values in Table 3 depict the cluster quality for 3 iterations. Figure 7 shows Silhouette plot for four clusters, value 0.9 indicates a good cluster quality with respect to cluster compactness and separation between the clusters. Figure 8 shows the dendrogram of 49 resumes. The results for different sample sizes are shown in Table 8 with important step of synset grouping and without synset grouping step. The resulted entropy shows significant improvement when synsets are used with grouping measure. The use of meronyms also shows improvement as resumes contain location data that forms concept hierarchy. There is also significant reduction in dimension and it remains almost steady with increase in sample size as indicated in Table 4. Figure 5 and 6 show dimension and entropy trend over number of resumes respectively. Table 5 shows the clusters obtained after processing 600 resumes and their features. Steps of Hierarchical approach with synset grouping are as follows

1. Apply cosine distance and hierarchical clustering algorithm to get document clusters.
2. Calculate Entropy to validate cluster quality
3. Extracting features of clusters at lowermost level
4. Extending feature set by adding cluster specific features
5. Repeating 1-6 till entropy stabilizes.

Table 3: Entropy readings for 49 resumes

|  |  |
| --- | --- |
| Iteration Number | Entropy |
| 1 | 0.22 |
| 2 | 0.18 |
| 3 | 0.06 |

Table 4: Comparative chart of Entropy values

|  |  |  |  |
| --- | --- | --- | --- |
| No of Resumes | Without Grouping of terms | Grouping using Synonyms | Grouping using Synonyms + Meronyms |
|  | Entropy | Dimensions | Entropy | Dimensions | Entropy | Dimensions |
| 49 | 0.44 | 56 | 0.27 | 45 | 0.032 | 35 |
| 200 | 0.53 | 70 | 0.31 | 55 | 0.041 | 40 |
| 600 | 0.64 | 90 | 0.33 | 65 | 0.052 | 42 |
| 800 | 0.67 | 95 | 0.36 | 70 | 0.061 | 42 |
| 1000 | 0.72 | 100 | 0.41 | 72 | 0.063 | 45 |

Table 5: Features for 600 resumes

|  |  |
| --- | --- |
| Cluster Number | Features |
| 1 | Enquiry, training, appraisal, strategy, mission, goal, objective, techniques, trend, application, support, maintenance. |
| 2 | Balance, audit, loan, finance, cash, asset, payroll, inventory, grid, Canada, Manitoba, Australia, Tasmania.. |
| 3 | Resume, planning, shipping, vendor, material goods, model, supply, culture, cross, claim, contract, conformance ,policy |
| 4 | Diploma, review, interpersonal, upgrade, conflicts,client… |

Figure 5: Dimension Plot.

Figure 6: Entropy trend over number of resumes.



Figure 7: Silhouette plot for four clusters



Figure 8: Dendrogram of 49 resumes

**Feature Extraction in quantified form**

Once the final clusters are formed, the final refined TDM gives the number of occurrences of each feature defined in the form of a synset group. The final feature matrix is obtained by normalizing the values to (0, 1) range. The feature matrix corresponding to a set of 24 resumes is shown in Table 6. Feature matrix for larger dataset of 1000 records where the entropy gets stabilized and the number of features is around 45 is shown in Table 7.

The experimental work establishes the efficacy of semantic based approach in getting the feature vector of each cluster and feature matrix in quantified form.

Table 6: Feature Matrix for 24 resumes

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | F1={Material,item, goods.} | F2={cash,Payment} | ….. | F19={Conformance, policy} | F 20={Australia, Tasmania} |
| R1 | 0.91 | 0.41 |  | 0.21 | 0.61 |
| R2 | 0.35 | 0.42 |  | 0.12 | 0.82 |
| R3 | 0.84 | 0.33 |  | 0.21 | 0.23 |
| R4 | 1 | 0.22 |  | 0.61 | 0.23 |
| ….. |  |  |  |  |  |
| R24 | 0.11 | 0.34 |  | 0.22 | 1 |

**Extracting features of skillset for a job position.**

The required skillset can be specified as a single model document giving all the required features or set of resumes that together describe the required skillset. In the later case iterative feature extraction can be applied to form clusters where each cluster represents the skillset and the feature matrix of skillset is obtained. The set of resumes provided can also undergo iterative feature extraction process to generate the feature matrix. Table 7 shows the cluster feature matrix where each cluster represents the skillset.

Table 7: Cluster feature matrix representing skillset

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | Plan | Support,Maintainance | ….. | Materials, goods | Oxford, England | Cumbria, Scotland |
| C1 | 1 | 0.41 |  | 0.31 | 0.18 | 0.41 |
| C2 | 0.12 | 0 |  | 0.12 | 0.45 | 0.23 |
| C3 | 0.84 | 0.42 |  | 0.71 | 0.21 | 0.12 |
| C4 | 0.76 | 0.91 |  | 0.61 | 1 | 0.87 |

The same process of iterative feature extraction is followed to extract features of new resumes. The intention is not to retrieve the clusters of resume, but to get the refined feature vector for each resume. Each feature vector describes the resume in the form of skillset. The skillset represented by each resume is shown in Table 8

Table 8: Features of Resume

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | objective | supply | … | audit | enquiry | conformance |
| NR1 | 0.34 | 0.12 |  | 0.87 | 0.41 | 0.11 |
| NR2 | 0.89 | 0.55 |  | 0.21 | 0.34 | 0.22 |
| .. | .. | .. | .. | .. | .. | .. |
| NR24 | 0.22 | 0.67 |  | 0.88 | 0.56 | 0.11 |
| NR25 | 0.11 | .76 |  | 0.18 | 0.87 | 0.78 |

**Recommending based on skillset mapping**

The features extracted from resume indicate the skill set of the user. The iterative feature extraction can be used to get refined features from a resume. The required skill set is represented by a set of model resumes. The iterative feature extraction can be applied to get the cluster feature vector where each cluster represents a required skillset. To select the correct resume that would fully satisfy the recruiter's need the mapping process is defined as follows

1. Identify feature vector(term set) for each skillset cluster and each resume using iterative feature extraction
2. Find cosine similarity between cluster feature vector and the skillset feature vector
3. The resumes with above threshold similarity will form recommended resume set for the skill set
4. The resumes having below threshold value for all skillsets will represent unmapped employee
5. The skillset having below threshold value for all the resumes will represent unmapped skillset

Feature set of cluster gives the parent terms of synonym and meronym groups. Table 9 shows the clusters of resumes and their terms. Some newly arrived resumes were then processed to extract their feature sets are represented in Table 11 and the cosine similarity measure distance was used to identify their proximity to cluster is shown in Table 10. The resumes in cluster can be recommended to the user as they have the features matching the user expectations. It was observed that the resume gets correctly classified in spite of the parent term’s absence in the resume.

Table 9: Feature set of clusters

|  |  |  |
| --- | --- | --- |
| Cluster | Terms | Resume assigned |
| 1 | Enquiry, training, appraisal, strategy, mission, goal, objective.. | NR2,NR3 |
| 2 | payroll, inventory, grid, Canada, Manitoba, | NR1,NR4,NR5 |
| 3 | model, supply, culture, cross, claim, contract, conformance ,policy | NR25,.. |
| 4 | Enquiry, training, shipping ,repair,… | NR6 |

Table 10: Cosine distance matrix showing new resume assignment to the cluster

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | C1(sales) | C2(accounts) | C3( Purchase) | C4(Customer service) |
| NR1 | 0.2312 | 0.91231 | 0.4532 | 0.3411 |
| NR2 | 0.8721 | 0.2321 | 0.3321 | 0.2212 |
| NR3 | 0.7621 | 0.1987 | 0.1784 | 0.1231 |
| NR4 | 0.3421 | 0.9834 | 0.4321 | 0.3412 |
| NR5 | 0.1234 | 0.8812 | 0.0091 | 0.2341 |
| NR6 | 0.3214 | 0.3412 | 0.1342 | 0.8765 |
| .. | .. | .. | .. | .. |
| NR25 | 0.3321 | 0.1891 | 0.9125 | 0.1341 |

**CONCLUSION**

The paper presents a task recommender system that uses semantic similarity between documents that can be used both for handling task allocation and also selection of new employees based on appraisal forms and resume. The most important pre requisite for this approach is the synset based feature extraction, semantic iterative clustering and semantic mapping. As Resumes may contain same content represented using different words and also contain a lot location specific and other concept hierarchies, both synonyms and meronyms were effectively used in forming the synset groups. The available dataset of resumes have been used but the system also needs to be validated on a real dataset of appraisal forms.

**REFERENCES**

1. Bafna, P.B., Shirwaikar, S. and Pramod, D., 2016. Multi-Step Iterative Algorithm for Feature Selection on Dynamic Documents. International Journal of Information Retrieval Research (IJIRR), 6(2), pp.24-40
2. Debnath, S., Ganguly, N. and Mitra, P., 2008, April. Feature weighting in content based recommendation system using social network analysis. In Proceedings of the 17th international conference on World Wide Web (pp. 1041-1042).ACM.
3. Derous, E., Pepermans, R., & Ryan, A. M. (2017). Ethnic discrimination during résumé screening: Interactive effects of applicants’ ethnic salience with job context. human relations, 70(7), 860-882.
4. Derous, Eva, and Ann Marie Ryan. "By any other name: Discrimination in resume screening." In The Oxford handbook of job loss and job search, pp. x-x. x, 2016.
5. Faridani, S., 2011, October. Using canonical correlation analysis for generalized sentiment analysis, product recommendation and search. In Proceedings of the fifth ACM conference on Recommender systems (pp. 355-358). ACM.
6. Glance, N., Meunier, J.L., Bernard, P. and Arregui, D., 2001, September. Collaborative document monitoring. In Proceedings of the 2001 International ACM SIGGROUP Conference on Supporting Group Work (pp. 171-178). ACM.
7. Gomez-Uribe, C. A., & Hunt, N. (2016). The netflix recommender system: Algorithms, business value, and innovation. ACM Transactions on Management Information Systems (TMIS), 6(4), 13
8. Halkidi, M., Batistakis, Y. and Vazirgiannis, M., 2002. Cluster validity methods: part I. ACM Sigmod Record, 31(2), pp.40-45.
9. Herlocker, Jonathan L., Joseph A. Konstan, Loren G. Terveen, and John T. Riedl. "Evaluating collaborative filtering recommender systems." ACM Transactions on Information Systems (TOIS) 22, no. 1 (2004): 5-53.
10. Horne, B. (2016). Resume Resume Service (RRS) Committee Planning Report.
11. Hung, L.V., Anh, N.T.K. and Dang, N.H., 2012, August. Improving Vietnamese web page clustering by combining neighbors' content and using iterative feature selection. In Proceedings of the Third Symposium on Information and Communication Technology (pp. 47-54). ACM.
12. Jannach, D., Zanker, M., Felfernig, A. and Friedrich, G., 2010. Recommender systems: an introduction. Cambridge University Press.
13. Jashki, A., Makki, M., Bagheri, E. and Ghorbani, A.A., 2009. An iterative hybrid filter-wrapper approach to feature selection for document clustering. In Proceedings of the 22nd Canadian Conference on Artificial Intelligence (AI’09) 2009.
14. Jayabharathy, J. and Kanmani, S., 2014. Correlated concept based dynamic document clustering algorithms for newsgroups and scientific literature. Decision Analytics, 1(1), p.1.
15. Karypis, G., 2001, October. Evaluation of item-based top-n recommendation algorithms. In Proceedings of the tenth international conference on Information and knowledge management (pp. 247-254). ACM.
16. Kille, B., Lommatzsch, A., Hopfgartner, F., Larson, M., & Brodt, T. (2017). CLEF 2017 NewsREEL Overview: Offline and Online Evaluation of Stream-based News Recommender Systems.
17. Kumar, A., Kumar, N., Hussain, M., Chaudhury, S. and Agarwal, S., 2014, December. Semantic clustering-based cross-domain recommendation. In Computational Intelligence and Data Mining (CIDM), 2014 IEEE Symposium on (pp. 137-141). IEEE.
18. Li, T., Zhang, Y. and Sindhwani, V., 2009, August. A non-negative matrix tri-factorization approach to sentiment classification with lexical prior knowledge. In Proceedings of the Joint Conference of the 47th Annual Meeting of the ACL and the 4th International Joint Conference on Natural Language Processing of the AFNLP: Volume 1-Volume 1 (pp. 244-252). Association for Computational Linguistics.
19. Luo, X., Zhou, M., Li, S., You, Z., Xia, Y., & Zhu, Q. (2016). A nonnegative latent factor model for large-scale sparse matrices in recommender systems via alternating direction method. IEEE transactions on neural networks and learning systems, 27(3), 579-592.
20. Melville, P., Gryc, W. and Lawrence, R.D., 2009, June. Sentiment analysis of blogs by combining lexical knowledge with text classification. In Proceedings of the 15th ACM SIGKDD international conference on Knowledge discovery and data mining (pp. 1275-1284). ACM.
21. Miller, B.N., Albert, I., Lam, S.K., Konstan, J.A. and Riedl, J., 2003, January. MovieLens unplugged: experiences with an occasionally connected recommender system. In Proceedings of the 8th international conference on Intelligent user interfaces (pp. 263-266). ACM.
22. Nunley, J. M., Pugh, A., Romero, N., & Seals, R. A. (2016). College major, internship experience, and employment opportunities: Estimates from a résumé audit. Labour Economics, 38, 37-46.
23. Pantel, P. and Lin, D., 2002, August. Document clustering with committees. In Proceedings of the 25th annual international ACM SIGIR conference on Research and development in information retrieval (pp. 199-206). ACM.
24. Rafi, M. and Shaikh, M.S., 2013. An improved semantic similarity measure for document clustering based on topic maps. arXiv preprint arXiv:1303.4087.
25. Risavy, Stephen D. "The Resume Research Literature: Where Have We Been and Where Should We Go Next?." Journal of Educational and Developmental Psychology 7, no. 1 (2017): 169.
26. Smith, B., & Linden, G. (2017). Two Decades of Recommender Systems at Amazon. com. IEEE Internet Computing, 21(3), 12-18.
27. Wei, T., Lu, Y., Chang, H., Zhou, Q. and Bao, X., 2015. A semantic approach for text clustering using WordNet and lexical chains. Expert Systems with Applications, 42(4), pp.2264-2275.
28. Zhang, F., Yuan, N. J., Lian, D., Xie, X., & Ma, W. Y. (2016, August). Collaborative knowledge base embedding for recommender systems. In Proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining (pp. 353-362). ACM.