Proactive System for Reviewer Paper Assignment

ABOLI PATIL[1], PARIKSHIT MAHALLE[2], VARSHA PATIL[3], SWATI BHAVSAR[4]
[1]Technology and Strategy, Munich, GERMANY
[2]Department of AI&DS, VIIT, Pune, INDIA
[3,4]Department of Computer Engineering, MCERC, Nashik, INDIA

Abstract: - In this paper, we propose a system that serves as a solution for the problem of automated assignment of reviewers to papers. With a steady increase in the number of research domains and huge submissions at journals and conferences, peer review happens to be the pivotal element to maintain quality standards for academic publications. Scientific and vigorous process for reviewer assignment is very crucial. Assigning appropriate reviewers poses a great challenge as it needs to consider many important aspects of like- relevance between reviewers and submissions, expertise, authority, diversity, recency and scientific impact. Existing approaches are based on matching the set of reviewers with submitted papers and assignment maximizes the similarity by satisfying the constraints such as load, coverage and conflict of interest. Traditional approaches are unsuccessful in i) identifying the multiple multi-disciplinary subject domains of paper and reviewer ii) assign a set of reviewers so as to cover all the subject domains of paper achieving higher topic coverage. The proposed system addresses both of these issues. The proposed is named as UPRPAS (Unsupervised Proactive Reviewer Paper Assignment System) uses Latent Dirichlet Allocation (LDA) based algorithm to build the topic model-based on the extracted contents of submissions and expertise of reviewers for calculating the similarity, and then find the best match and assignment. The basic idea is to inevitably build representations of semantically relevant aspects of both papers and reviewers in order to facilitate the construction of a relevance matrix. The performance of the proposed systems is evaluated using conference datasets and is compared with baseline algorithms. Experimental results show that paper and reviewer profiles are built more accurately with higher collective matching degree and topic coverage. The systems accurately perform the assignment of reviewers to papers. The work also contributes a reviewer matching dataset and evaluation that will be useful for further research in this field.

Key-Words: - Bag of Words (BoW), Coherence of Topic Model, Latent Dirichlet Allocation-LDA, Perplexity, Reviewer Assignment Problem-RAP

Received: April 25, 2024. Revised: January 14, 2025. Accepted: March 15, 2025. Published: May 13, 2025.

1 Introduction

Reviewer assignment is a process of assigning the most appropriate reviewers to the submitted papers for fair and accurate reviews. The problem of identifying appropriate experts for reviewing papers and their assignment to paper satisfying constraints is known as the Reviewer Assignment Problem (RAP). Constraints here mean that the assignment of the reviewers to the papers must satisfy a few conditions for the fair and accurate review process. A typical set of constraints include- load, coverage, and conflict of interest. Process of assignment is to be done in a way that: Every paper should be evaluated by a minimum count of reviewers called coverage, Load is the number of maximum papers to be assigned to the reviewer and Reviewers assigned to a paper should not fall into a conflict of interests.

One of the most popular applications of reviewer assignment is in the conference and journal management systems. For the last two decades, the conferences and journals are loaded with excessively high numbers of paper submissions. An important task of the editor or conference chair is to get the papers reviewed accurately and within the deadline. Publication of research papers or articles at journals and conferences has been very close to the heart of all researchers and academicians. One of the key components associated here is the review process that includes the task of assignment of the reviewers to the submitted papers. The submitted papers are also referred to as manuscripts. Peer review is defined as a process to validate research and academic work, helping to improve the quality of published work and develop strong networking within research communities. This task of assignment of papers to reviewers is the most crucial and challenging. It includes the highly important task of identifying the appropriate and competent experts in the paper-specific domain from maybe registered reviewers. Figure 1 indicates the typical process of submissions of papers, their scrutiny, and assignment to the reviewers.

The study reveals that the acceptance rate at various journals of repute is very low. The acceptance rate depends on the quality of the submitted paper and the quality policies defined. These rates contribute to the quality control standard internally, and the impact factor of the journal contributes to external standards. It is also apparent that the quality of the reviews hampers the reputation of conferences and journals. Most of the researchers feel as if they are trapped in a process of submission, reviews, rejection, revision, resubmit and re-review and so on till re-re-review [2, 3, 4]. This seems to eat up months of the researcher's life, hampers career, funds, rage, and delays in the dissemination of results. The study exhibits a very strong correlation between the review process, reviewer selection, and review quality [9, 10]. An investigative study observed the impacts of peer assessment as a social affective parameter and students' displeasure related to the peer assessment. And it is also observed that the accuracy in the process of reviewer assignment augments the process of peer review [2, 10, 15]. The accurate selection of the reviewers surely guards the quality of the process against reviewer rude behavior. Further researchers positively acknowledge the thumb rule - 'to seek reviews by multiple reviewers covering most of the paper domains, and it is noticed that most of the journal editors and conference chairs use this thumb rule. To address all these issues, the reviewer assignment problem has attracted researchers' interest in recent years [8, 12]. In the early days, reviewer assignment was performed manually by editors or conference chairs. However, with the increasing number of academic exchanges, the number of reviewers and papers has risen exponentially. So the manual process of reviewer assignment is not feasible and is prone to errors. Researchers claim that an automatic computer-assisted reviewer to paper assignment system can serve the purpose [14]. The study reveals that there is an urgent need to replace the manual system as it is complex, infeasible, and error-prone due to the high count of papers and most of them cover multiple subject domains. The recent research in the RAP domain focuses on the automation of the whole process with improved accuracy [10, 15]. Along with journal and conference, reviewer assignment is a vital task in many research activities like evaluating grant proposals, teacher class assignments, course examiner assignments and many others.

UPRPAS is an unsupervised Latent Dirichlet Allocation based technique for building a topic model for the papers and reviewers. For each paper and reviewer, multiple subject domains are accurately identified and relevant labels are generated. The outcome is annotations for papers and reviewers dataset; treating these annotations as labels, serves as input for supervised learning. An algorithm for similarity matching and ranking is developed that generates topic lists, relationships among topics, and topic dictionaries. Further, it computes relevance and performs the ranking of reviewers based on a number of publications in the topic domain, recency, and h-index. The basic idea is to inevitably build representations of semantically relevant aspects of both papers and reviewers in order to facilitate the construction of a relevance matrix. An apparent choice of such a representation for papers and reviewers' publications are as a weighted bag-of-words that is the set of distinct terms in documents D, vocabulary V, defines a vector space with dimensionality |V| and thus each document d is represented as a vector in this space. The query q can also be represented as a vector in this space, assuming it shares vocabulary V. The query and a document are considered similar if the angle q between their vectors is small. The angle can be conveniently captured by its cosine, giving rise to the cosine similarity. The study and experimentation revealed that the accuracy of reviewer paper assignment depends on the accuracy of the key phases that work in sequence- identifying the subject domains, computation of similarity, matching, and assignment [1, 2, 15]. Errors in one phase may get propagated to the next phase and similarly, the accuracy in each phase surely contributes to the improvement of the overall accuracy of the system. Keeping this in mind, the proactive approach is used for the proposed system development and implementation. An attempt is made to proactively identify the parameters that contribute to accuracy and take appropriate measures to prevent the errors that have been anticipated.

The rest of the paper is organized as follows. Section 2 describes the reviewer assignment problem. In Section 3, the Reviewer

Assignment Process is explained. Section 4 describes the status of research work in the Reviewer Assignment Problem with potential research gaps and challenges, section 5 presents proposed systems. Section 6 presents experiments carried out and results, including details of the primary dataset we created for our mainline methodology, comparisons against alternative approaches and choices, and indirect evaluation on an available prior dataset. Finally, Section 7 concludes the paper with a summary of the main findings and a discussion.

2 The Reviewer Assignment Problem

The process of reviewer assignment to paper is a typical example of a classical optimization task where some constraints are to be satisfied and there are limited resources like reviewers. For a given set of papers and a set of reviewers, The objective is to assign the most appropriate reviewers to a paper. The objective is that the assignments of the papers to the reviewers should be made so that the total matching degree is maximized (equation 1). The expected outcome is assignment of the paper to the reviewer with the high relevance and the low conflict of interest satisfying the constraints of load and coverage (equation 2, 3, 4).

Let R = {ri} for i=1 to N be the set of reviewers,

P = {pj} for j=1 to M be the set of papers and

A $\in \mathcal{R}|R|\times|P|$ be a matrix of reviewer-paper similarities also known as an affinity matrix.

Given R reviewers and P papers reviewer assignment problem is expressed as:

$$\text{Max} \sum_{i=1}^{|R|} \sum_{j=1}^{|P|} xji\ Aij \qquad \ldots\ldots\ldots\ldots\ldots (1)$$

subject to

$$\sum_{i=1}^{|R|} xji \ \leq Ui \ \ \forall\, i = 1, 2, ..., |R| \ \ldots\ldots\ldots\ldots\ldots (2)$$

$$\sum_{j=1}^{|P|} Aji \ \leq Ci \ \ \forall\, j = 1, 2, ..., |P| \ \ldots\ldots\ldots\ldots\ldots (3)$$

$$\text{CoI (Pi, Rj)=False} \qquad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots (4)$$

xji Є {0,1} and {Ui} is the set of upper bounds on reviewer loads, and {Cj} represents the coverage constraints, and CoI is conflict of interest.

The matching of reviewers to papers is encoded in the variables x,

If xji is set to 1then it indicates that reviewer ri has been assigned to paper pj ,

{Ui} is the set of upper bounds on reviewer loads,

{Cj} represents the coverage constraints, and CoI is conflict of interest.

In this formulation, the objective is to maximize the sum of affinities of reviewer-paper assignments subject to the listed constraints.

3 The Reviewer Assignment Process

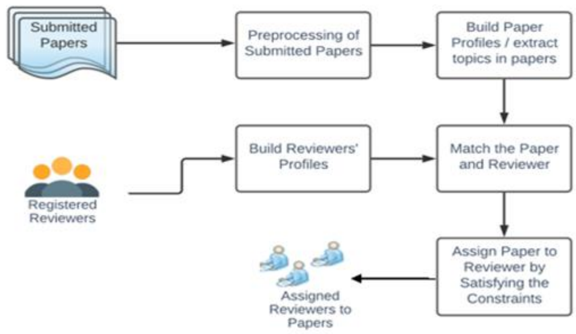

Fig.1: The typical process of the reviewer to the paper assignment

Most of the researchers have divided the reviewer assignment process into four major steps. Input is a list of submitted papers and a set of identified or registered reviewers. After pre-processing of the manuscripts, typically subject domains of papers are extracted and similarly, the expertise of reviewers is identified. Lastly, the matching and assignment are done. Core Phases are - Paper Profile Building, Reviewer Profile Building, Affinity Matrix Computation, Matching and Assignment.

Paper Profile Building

The first phase is paper Topic Modeling that is also named paper profile building. It is the process of finding the subject domain/topics of paper using paper contents.

Reviewer Profile Building

The second phase is computing expertise of reviewer and is also named as reviewer profile building that is done with existing publications of expert, registered information such as choice of tracks, confidence level and collected information such as a number of publications, citations, designation and similar.

Computation of Affinity Matrix

The third phase is to compute the similarity between paper and expert using the profiles built.

Match and Assign the Reviewers to Paper

The last phase is to match reviewers and papers using an affinity matrix and from the set of matching experts, assign the most appropriate reviewers to the paper satisfying the constraints.

Constraints

Typical set of constraints for the reviewer assignment problem include-

Load- Maximum number of papers per reviewer,

Coverage- Minimum number of reviewers per paper,

Conflict of Interest (COI)- Assignment of reviewers avoiding conflict of interest

Along with typical constraints load, coverage and COI, topic coverage is an important constraint that needs to be satisfied. Recent decade has witnessed the submission of a huge count of papers that are multidisciplinary and interdisciplinary covering more than one subject domains. The reviewer assignment must assure that for each paper, the coverage of paper domains by the collective expertise of assigned reviewers is maximized. The quality of reviews is measured with topics covered collectively by the expertise of assigned reviewers.

4 Status of Research Work in Reviewer Assignment Problem

Total of 180 research papers from journals and conferences confined to the reviewer assignment problem and related topics have been studied to know the status of the research domain to understand the potential research gaps and challenges.

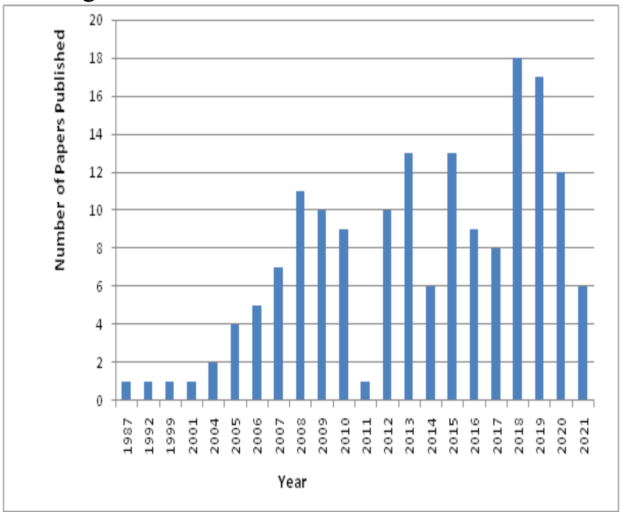

Fig.2: Yearly Publications Related to Reviewer Assignment Problem

Figure 2 shows the yearly publications related to RAP and related subjects indicating the noticeable count of research papers. We can categorize this publication as Peer Review process and related issues, Building of reviewer Profile, Building of Paper Profile, Reviewer Assignment Problem, Topic Matching and Ranking and Survey Papers.

4.1 Potential Research Gaps and Challenges

It is observed that even though the submitted papers at conferences and journals are covering multiple domains and need experts with knowledge of more than one subject domain; researchers haven't identified more than one topic for both papers and reviewers. Expert's recent publications in the paper-specific domain mirror the knowledge and research interest directly, and citations help to understand his/her authority and mutual recognition in the academic area, but the study reveals that it is not utilized well. Most often based on reviewers selected track/domains the papers are assigned.

Paper profile building utilizes only two sections of manuscript as-Title and abstract leading to inability to accurately identify the paper topic domains. Indeed sections like introduction and conclusions can help in more accurately identifying the paper topics. Further, the topic extraction technique processes the papers and experts' publications separately. This hampers the consistency of topics and processing them together may lead to better topic modelling.

Proper policy for ranking the selected reviewers based on a number of publications in the paper-specific domain, on recent publications in the paper-specific domain, and on h-index, and citations assuring the expertise and authority to review is missing.

The dataset of the final assignment of reviewers and their expertise remains with the respective conference organizing team and is neither disclosed nor is made available for others to use. An accurately labeled set of papers with assigned reviewers is not available. For supervised learning techniques, labeled data is required. Also, for measuring the performance of reviewer paper assignment techniques, a reference dataset with output is necessary. Most of the techniques involve human assistance. It is hard to accurately build a data set that is labeled manually even though editors with sufficient experience are invited to do so.

The key challenge is how to notify the machine regarding the field knowledge of the expert. Challenge is how to train the machine to gain knowledge. Experts have wide-ranging research interests in various fields. Challenge is to utilize the diversities in reviewers' research interests explicitly by capturing their expertise comprehensively. High topic coverage is challenging as it assures collective coverage of all topic domains of the paper by the cumulative expertise of reviewers. We have very well addressed this challenge by treating the Topic Coverage as the main constraint and satisfying it. The output is a dataset with the reviewer-paper assignment that is made available for performance evaluation and Comparison by further researchers.

5 Proactive System for Reviewer Paper Assignment PRAPS

The proposed system aims to provide a machine learning specially an unsupervised learning approach based solution for reviewer assignment problem. Following is the rationale behind the proposed system design. The proactive novel system is designed as a solution for the Reviewer Assignment Problem to proactively identify the parameters that contribute to accuracy and take appropriate measures to prevent the errors that have been anticipated. Proper ranking of the matched reviewers based on recent publications in paper specific domains, and impact (based on h-index, and citations) adds to the accuracy.

5.1 Rationale

In addition to title and abstract; keywords, introduction and conclusion can extract topics more accurately. Topic coverage utilizing diversity of expertise of reviewers assuring that at least one reviewer per topic will assure the appropriate reviews.

The accuracy of reviewer paper assignment depends on the accuracy of all the core phases in cascade- building profiles, computation of similarity and matching and assignment. Errors in one phase may get propagated to the next phase and similarly, the accuracy in one phase will surely contribute to improving the overall system accuracy. Keeping this in mind, the proactive approach is followed for the proposed system development and implementation. An attempt is made to proactively identify the parameters that contribute to accuracy and take appropriate measures to prevent the errors that have been anticipated. The key goal is to make accurate assignments of reviewers to submitted papers for fair and accurate reviews. Keeping this in mind, the proactive approach is used for the proposed system development and implementation.

5.2 Dataset & Experimental Setup

The input datasets used for experimentation are prepared manually by downloading pdfs and extracting the required sections using four conference papers AAAI 2020- 100 papers, AAAI 2019 – 398 papers, NIPS 2014-1425 papers, and Interspeech 2019- 145 papers; a total of 2068 papers. These conferences are preferred as the proceedings of conferences with full paper pdfs are made available for researchers to refer to as open access. The dataset for 106 reviewers is prepared by collecting their 792 publications from Google scholar and extracting required sections from pdfs. Fields used are ID, Name, Affiliation, number of publications, h-index, i-10 index, citations count. The data are collected from academic resources such as the DBLP Computer Science Bibliography, ResearchGate, and CiteSeer, which are available in public domains.

All experiments are conducted on the system Intel i5@2.50GHz with 8 GB of memory. Proposed methodologies are implemented using python with the help of various NLP and machine learning libraries (nltk, scikit-learn, keras, pandas, tensorflow, gensim etc). Topic models for each dataset are built distinctly by keeping LDA hyper-parameters constant and varying the number of topics. Constraints: number of reviewers assigned to the manuscript (c) is set to 5 and the Maximum number of manuscripts assigned to a reviewer (m) is set to 10. Top 5 relevant topics obtained for manuscripts and reviewers' publications from topic distribution are used to calculate manuscript to reviewer relevance.

5.3 Architecture of Unsupervised Proactive Reviewer Paper Assignment System (UPRPAS)

The proposed UPRPAS comprises of core seven functional process blocks as shown in figure 3. The core blocks of the proposed UPRPAS are as follows-

1. Pre-processing and Corpus Building
2. Topic Modeling and Label Generation
3. Paper Profile Building
4. Reviewer Profile Building
5. Proficiency Computation
6. Reviewer Ranking
7. Reviewer to Paper Assignment

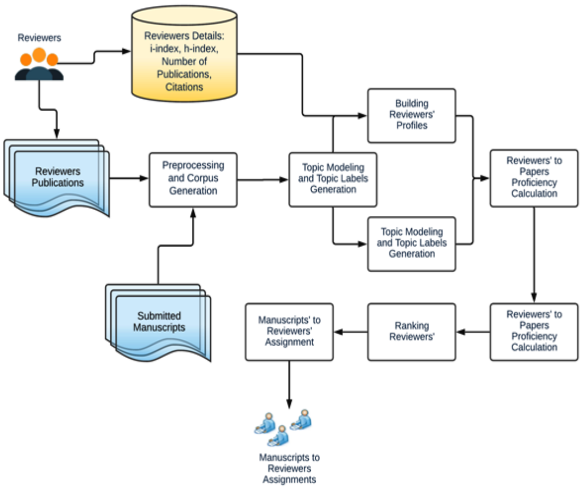

Fig.3: The architecture of proposed system-UPRPAS

5.3.1 Preprocessing and Corpus Building

The input to the system is a text corpus that is built using text extracted from various sections of submitted papers. Traditionally researchers have used only the title, abstract, and keywords and the proposed system included introduction and for some data sets conclusions section too. This inclusion of introduction and conclusion sections of papers in the corpus has improved the topic coverage of the topic model. The introduction and conclusion sections capture the field domain and latent topics in the paper that may not be included in sections title and abstract as they are too precise and short. Experimental results indicate that better clustering results can be yielded with higher sized corpus by accurately differentiating and co-relating topics. One single corpus is built with both submitted papers and publications' of reviewers that leads to a reduction in dimensions of the profiles. The collected text corpus is pre-processed before topic modeling so as to prepare the text for use in topic modeling and analysis. A Series of steps like clean and normalize are followed in text processing. Popularly, text cleaning, text tokenization, special characters removal, conversion of case, spell correction, stop words removal, stemming, and lemmatization are performed. The output of pre-processing is normalized corpus.

5.3.2 Topic Modeling and Topic Labels Generation

One of the most important tasks is to build reviewers' and manuscripts' profiles. Building reviewers and paper profiles need to extract the topic distribution across manuscripts and reviewers' publications. Topic modeling and key phrases extraction techniques are used for the same. Topic modeling is basically a mathematical and statistical modeling method which extracts foremost topics or ideas from a corpus of documents. The concept of the topic model is to extract the important fields or perceptions from a corpus built with papers and represent them as topics. Individual topics are represented using a set of terms from the corpus. Collectively, these terms imply a precise topic and each topic can be easily differentiated from other topics by understating and analyzing the semantic meaning carried by these terms.

The proposed methodology uses a Latent Dirichlet Allocation (LDA) to build the topic model which is a probabilistic statistical approach [5, 6, 11]. Implementation of the LDA model is discussed in detail in the forthcoming section. Topic modeling and key phrases extraction together is used to extract key research topics and description of each topic. Set of key phrases extraction is the process of identifying phrases from a corpus to capture the core subject domains. The topic here refers to a set of words defining a specific field/domain/area, for example, for the topic 'education', the set of words like a teacher, student, classroom, books, and similar words direct us to identify the domain. It is an unsupervised learning technique that infers the latent topics from a provided corpus of documents that are submitted papers. Each paper represents a distribution of identified topics while each topic is a distribution of words (or phrases). The key expertise domains of reviewers are represented as probability distributions on more than one domain.

5.3.2.1 Feature Engineering

Traditional TF-IDF, Bag of Words (BoW) models, and Bag of N-Grams models are inherent in nature and they are just bags of words. These models are not capable of extracting semantic structure, text sequence, and context around neighboring words in the document. In the proposed work experimentation is done to overcome these drawbacks. After pre-processing on texts (titles + abstracts + keywords + introduction + conclusion) from manuscripts and reviewers' publications output is a collection of vectors of tokens V for each manuscript and reviewer publication collectively. It is required to transform the textual data into a machine-understandable form as the machine doesn't understand the text. Feature engineering is a vital step; it aims towards transforming unstructured, textual data into numeric representations which then can be fed as input to machine learning algorithms. Each paper is represented as a Bag of Words model by a numeric vector having dimension represented with a specific word extracted from the corpus with value representing its frequency in the paper.

Before vectorization, n-gram based important phrases are extracted from the cleaned corpus (V) and unnecessary terms are removed. As a phrase or sentence reveals more semantic details than a single word, initially, extraction and generation of bi-grams and tri-grams as phrases is done. For this, a phrase extraction model on corpus V is built. The min count parameter which serves (μ), is used, which states that the phrase model ignores all terms and bi-grams with a total collected count lower than (μ) across the corpus (V). The value of (μ) varied from 2-10 during experimentation. Tri-gram phrases (Ft) are generated for each research paper and reviewer's manuscripts by applying the phrase model.

Fb = Mp(V) ………………(5)

Where Fb = bi-gram phrases

Ft = Mp(Fb) …………….(6)

Finally, vocabulary (D) is generated from corpus V . D is a dictionary representation of phrases in the corpus, which is unique phrase to number mapping. D = {(n1, f1), (n2, f2), ...}, where fi ∈ F and ni is number mapped to fi. Equations (6, 7) summarizes the corpus to vocabulary generation process.

D = BoW(Ft)…………….(7)

5.3.2.2 Building Topic Model

The proposed approach uses the Latent Dirichlet Allocation (LDA) algorithm to extract the topics covered in manuscripts and reviewers' publications which are further used in reviewers' and manuscripts' profile building and to calculate relevance between them. Use of LDA and proposed approach overcomes the challenges of semantic mismatch and the computational complexity. The two key hypotheses of LDA BoW (bag of words) and BoD (bag of document) serve as base for three-level hierarchical (document_topic_word) Bayesian model. The author David et al. assures that "The fundamental principle is that documents are characterized as infinite mixtures over latent topics, where each topic is represented as a distribution of words" [4,7]. LDA represents each topic as the distribution of words belonging to it. Let us consider two topics say named 'Machine Learning' and 'Graph Partitioning' Words like 'training', 'neural', 'epoch', 'over-fitting' and similar may have a higher probability distribution for the topic of machine learning over words like nodes, edges, cut set, and similar. On the other hand, words like nodes, edges, cut-set, partitioning, visualization, and similar may have a higher probability distribution for the topic graph partitioning over words like over-fitting, training, epoch, and similar. Each topic may share a similar group of words with a higher probability.

Algorithm 1: Algorithm for Topic Model Building
Input: Set of papers P, Set of Parameters
Output: Topic distribution within corpus

1. *Initialize all required parameters.*
2. *For each papers, p in P:*
   a. *For each phrase/word f in p:*
      i. *randomly initialize each word to one of the **K** topics*
3. *For each iteration:*
   a. *For each paper, p in P:*
   b. *For each phrase/word in p:*
      i. *For each topic T in K*
         1. *Calculate P(T/p), proportion of words in d assigned to topic T*
         2. *Calculate P(f/T), proportion of words assignments to topic T over all paper having phrase f.*
         3. *Reassign phrase f with topic T with probability P(T/p) x P(f/T), considering all other phrases and their topic assignments*

Algorithm 1 describes the working of topic model building steps. Topic modeling is done with an unsupervised approach based on LDA. LDA assumes that documents with similar topics use a similar group of words. This enables the documents to map the probability distribution over latent topics. While building the LDA model the first important challenge is to decide the number of topics, say k to generate and initialization of other parameters discussed earlier. The LDA model is built by varying values of k by keeping the other parameters constant. Experimentation was conducted and the different values for k and Other parameters are kept as it is to their default values. Steps in the algorithm were run empirically for 500 iterations to build the LDA model, which outputs the topic mixtures for each document and then constituents of each topic from the terms that point to that topic obtained.

5.3.2.3 Optimizing Number of Topics

The discovery of the optimal number of topics in a topic model is challenging, and it is required to set before training the model. With an iterative approach and with the numerous models built by

varying numbers of topics, the model is selected that has the highest coherence score (Cv).
After verification, the finalized number of topics with the coherence score, number of different topics for NIPS2019 is shown in table 1. Number of topics is set as K=20 after rigorous experimentation.

Table 1: Topic Models Coherence Score and Number of Topics

Number of Topics	Coherence Score
5	0.31192093612615385
7	0.3450840709698132
10	0.36712104133461
12	0.38563331276218693
15	0.4091361272818827
18	0.41313022976112107
20	0.4251539103846886
25	0.4304065707837326

The optimal number of topics, K = 20 are chosen. Topics obtained and term distribution for each topic is as shown below, which visualize the topics as tuples of terms(X) and weight of each term for respective topic.
X = {xt1, xt2, ..., xtK} ……………(8)
where, xti is the set of terms generated for topic ti.
xti = {f1,f2, f3, ...} ……………(9)
where fi is a term or phrase and weight of it for topic.

Topics obtained are easy to understand and represent the importance of each term in the topic. To evaluate and to measure the quality of the topic model, the mean coherence score (Cv) and perplexity (Pv) of the topic model are calculated. Classically a set of statements is considered to be coherent if they support each other. Typically, when the perplexity is lower and mean coherence score is higher, then the model is said to be better.

Number of Iterations
The number of iterations is decided by comparing the training perplexity (Pv) of the LDA model on the whole corpus under a different number of iterations, as shown in figure 4. It is noticed that at 500 iterations perplexity tends to be stable and the value for after experiments for number iterations is set as 500.

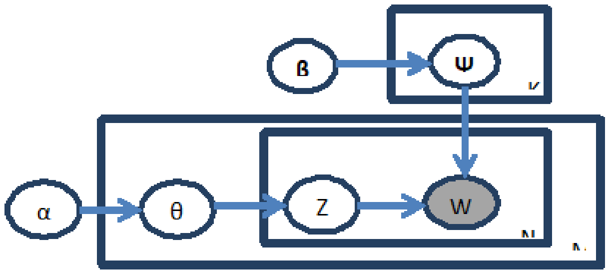

Fig. 4: LDA (Mathematical Model)

Number of Top Terms
The top terms (k) value is selected as 200 in the order of descending weight to represent each topic. This number is chosen empirically, which covers approximately 70-80 percent of the probability space of each topic.

5.4 Building Paper Profile

Building a paper profile is an essential task aiming to extract the key topics in the paper. Building a manuscript profile is an essential task in peer review, which aims to find the topic coverage in the manuscript and extract the key topics in the manuscript. Topic distribution in the manuscript is obtained by applying the LDA model, which gives the weight of all topics in T to manuscripts P, WP ,T
.

WP ,T = {{wp1,t1,wp1,t2, ...,wp1,tk}, {wp2,t1,wp2,t2, ...,wp2,tk, ...,wp|p|,t1,wp|p|,t2, ...,wp|p|,tk}} *(10)*

where pi is ith manuscript and tj is jth topic. Document to topic weight signifies the relevance between topic and document. Higher is the weight of a document to the topic which means the document is more relevant to the topic. The primary goal is to obtain the key topics covered in each manuscript, pi ∈ P and each reviewer's publications, qi ∈ Q. This helps to understand the key research fields covered in each manuscript and discover the reviewers having expertise in these _elds. In order to do the multidimensional analysis _rst, it is required to decide how many topics are to be considered as relevant (η), among research topics covered in each paper. Further η number of most relevant topics for each manuscript (Γpi) are obtained. Here the value of η to 5 is empirically set. Out of K topics, 5 topics (Γpi) extracted which are having maximum weight for manuscript pi. In the same way, the most relevant topics in all manuscripts P are extracted.

Γ = {Γp1, Γp2, ..., Γp|P|} ……………… *(11)*

where, Γpi is collection of most relevant topics for manuscript pi.

Γpi = {(t1,wpi,t1), (t2,wpi,t2), ...(tη,wpi,tη)},where, η = 5 ……(12)

For manuscript pi, key topics Γpi are arranged in descending order by their weights, it simply means that topic ti is most relevant than topic ti + 1 to manuscript as shown in algorithm 2.

Algorithm 2 : Manuscript profile building

Input: $P = \{p_1, p_2, ..., p_{|P|}\}$*; manuscripts*
η = 5; number of topics considered as most relevant to manuscripts
Output: Γ = collection of most relevant topics for manuscripts

```
1.  start
2.  Γ = []
3.    for each manuscript p in P
4.        Wp,T  = LDA(p)
5.        Γp = []
6.        for i =0 to η-1
               Γp[] = (ti, W_{p, ti})
7.  Γ[] = Γp[]
8.  End
```

5.5 Building Experts Profile

Similar to a paper profile, a profile of experts is built. Building a reviewer profile aims to find the most relevant reviewer. A person with maximum match between expertise with a paper topic and having recent publications in a paper topic domain with higher impact is the most relevant reviewer for the paper. Reviewers' profile is built using the publications of him or her. The publications of experts help us to know the expertise in subject domains. The Citations, and h-index help in measuring the authority of the reviewer for a particular topic. h-index, or Hirsch index, measures the impact of a particular scientist rather than a journal. "It is defined as the highest number of publications of a scientist that received h or more citations each while the other publications have not more than h citations each. It helps to measure the impact of his expertise. And of course among many publications, the subjects having recent publications assure his recency in expertise. With this underlying principle, We have built the reviewers profile with Expertise, Authority and Recency.

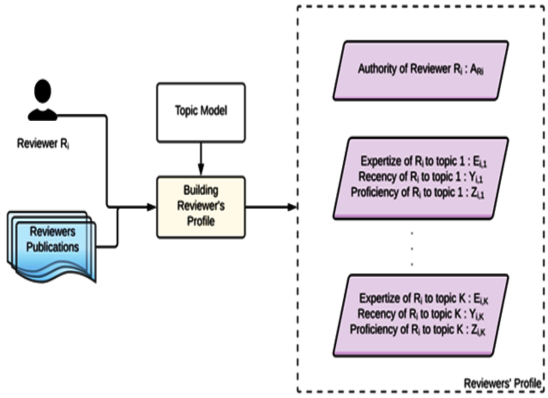

Fig.5: Reviewer Profile Building

5.5.1 Expertise

For peer review, it is the default expectation that reviewers should possess expertise in the fields or the topics covered in the manuscript. Reviewer's expertise can be obtained in various ways like each reviewer can choose or mention his or her research areas, reviewer's areas of interest can be obtained from various academic data sets and platforms like Google Scholar, DBLP and similar. In the proposed approach, topic modelling is used to automatically extract key topics from reviewer's publications to gain the expertise of reviewer. To extract reviewer expertise from his or her publications, all reviewers' publications' documents are collected, say Rp, from Google Scholar for all reviewers R.

To compute the expertise of a reviewer ri first we need to extract the topic coverage from the reviewer's publications (qi). The LDA model built earlier is applied on each publication qi,j , where qi,j is jth publication of ith reviewer. LDA module outputs the topic distributions for K topics, publications to topics weights and term distribution in each publication.

Each reviewer to topic weights obtained by considering all publications of reviewers. From each publication of reviewer title, abstract, keywords, and introduction are combined to form a single document which collectively represents all publications of a reviewer as whole (qdi).

$$q_{di} = \sum_{j=1}^{|q_i|}\Big(title_{q_{i,j}} + abstract_{q_{i,j}} + keywords_{q_{i,j}} + introduction_{q_{i,j}} + conclusion_{q_{i,j}}\Big) \ldots(13)$$

q_{di} to topics weights $W_{qdi,T}$ applied by applying LDA model on $q_{di,}$

$$W_{qdi,T} = \{W_{qdi,t1}, W_{qdi,t2}, \ldots\ldots, W_{qdi,t|K|}\} \ldots\ldots\ldots\ldots\ldots\ldots\ldots(14)$$

Similarly, reviewer to topics weights obtained for all reviewers in R, which forms a reviewers to topics weight matrix WR,T.

As most researchers work in multiple domains, so while building a reviewer profile it is required to decide a number of topics for which reviewer is most relevant, say V. Here empirically the value of v is set to 5. Expertise is calculation of topic relevance or extracting the expertise of a person. To obtain the most relevant topics to each ith reviewer say Fri, reviewer to topic weights is sorted in descending order, and then out of K topics first v topics are considered as most relevant topics for a reviewer. Here, Ei,j is the expertise level of ith reviewer for topic ti and wt is the weight of reviewer for topic t and F represents the expertise matrix for all reviewers for all topics.

$E_{i,j} = (t_i, w_{ri}, t_i)$

$$F = \{F_{r1}, F_{r2}, \ldots F_{r|R|}\} \ldots\ldots\ldots\ldots\ldots\ldots(15)$$

$$F_{ri} = \{E_{i,1}, E_{i,2}, \ldots, E_{i,\upsilon}\} \ldots\ldots\ldots\ldots\ldots\ldots(16)$$

Where, Ei, j = (tj, wri,tj), is expertise level of ith reviewer for topic tj and wt is weight of reviewer for topic t.

Here F represents the expertise matrix for all reviewers. Sample expertise matrix is shown in table 3.7. Weight of ith reviewer for jth topic is termed as expertise of reviewer for topic (Ei,j).

5.5.2 Authority

The authority of a reviewer (Ari) indicates the quality and quantity of publications. It indicates experts' recent status of research domain and expertise in the domain. The authority of the expert is computed using parameters number of publications, h-index, number of citations collected from the Google Scholar. Here authority of ith reviewer Ari is computed as one by one plus e raise to ¼ of number of publications plus h-index plus number of citations divided by 4. Here we used a sigmoid function to range the value of authority between 0 to 1. Usually the number of citations of authors is higher than the number of publications, thus we have considered 4 citations per publication and added 1/4 as constant to normalize the competence function.

$$A_{ri} = \frac{1}{1+e^{1/4(Q_{Nri}+H_{ri}+I_{ri}+\frac{C_{ri}}{4})}} \quad \ldots\ldots\ldots(17)$$

The authority of the expert is computed using parameters number of publications (Qnri), h-index (Hri), number of citations (Cri).

5.5.3 Recency

The reviewer should be active recently in research areas covered in the manuscript. It is required to analyze the complete research career of the reviewer and if there is a deficiency of recency then full recognition is not given. The recency of reviewers for the topic (Yri, tj) is measured by considering the span of Y years and most relevant publications to the topic (qri, Tj). Here the span of 10 years is considered. In order to obtain the recency of the reviewer, the first recency of each publication of the reviewer (Yqi,j) is calculated using equation 18 and 19.

$$Y_{qi,j} = 10 \times (\text{current\_year} - \text{publication\_year of } qi, j) \ldots\ldots\ldots(18)$$

The recency of ith reviewer for jth topic is defined as follows in equation 19.

$$Y_{i,j} = \frac{1}{10\,|q_{ri,Tj}|}\sum_{q \text{ in}}^{q_{ri,Tj}} Y_q \quad \ldots\ldots\ldots\ldots\ldots(19)$$

Recency of ith reviewer for jth topic is Yi,j Using equation 19, recency of all reviewers for K topics is calculated. The recency value obtained is normalized between -1 and 1.

5.5.4 Proficiency

A novel algorithm and term proficiency is introduced that represents a value computed with a reviewer's expertise, authority and recency. This algorithm has improved system efficiency by cutting the time complexity.

The proficiency of the reviewer represents a value computed with a reviewer's expertise and relevance to each topic in K. This is further used to measure the relevance between reviewer and manuscript. Proficiency of ith reviewer for jth the topic is calculated using reviewer to topic expertise (Ei,j), reviewer authority (Ari), and reviewer to topic recency (Yi,j). The proficiency of a reviewer for a topic (Zi,j) is defined as follows:

$$Z_{i,j} = \alpha\, E_{i,j} + \beta\, Y_{ri} + \gamma\, A_{i,j} \quad \ldots\ldots\ldots\ldots\ldots\ldots(20)$$

Here in equation, E is expertise of reviewer representing reviewer to topic_i relevance. R represents topic_i recency of expert computed using each topic recency. A represents authority of expert computed using experts' total publications, citations, and h-index Proficiency of ith reviewer for jth the topic (Zi,j) is calculated using reviewer to topic expertise (Ei,j), reviewer authority (Ari), and reviewer to topic recency (Yi,j). α = 0.5, β = 0.3, and γ =0.2 are concentration parameters which signifies the importance of Ei,j , Yi,j , and Ai,j respectively.

Algorithm 3: Ranking Algorithm

Input: $R = \{r_1, r_2, ..., r_{|R|}\}$*; set of Reviewers*
$Q = \{q_{1,1}, q_{1,2}, q_{1,3}, ... q_{5,1}, ... q_{10,1}.........\}$*, where* $q_{i,j}$ *denotes* j^{th} *publication of* i^{th} *reviewer,*
Reviewer details,
LDA model,
υ *= number of topics to which reviewer is most relevant,*
Output: Reviewers' profiles
A =Reviewers' authorities
F = Reviewers' expertise matrix
Y = Reviewers' recency matrix
Z = Reviewers' proficiency
1. *start*
2. $F = [], Y = [], A =[]\ Z = []$
3. *for each reviewer r in R*

$$A_r = \frac{1}{1 + e^{1/4(Q_{Nri}+H_{ri}+I_{ri}+\frac{C_{ri}}{4})}}$$

$A[] = A_r$
$F_r = []$
for each publication q of r

$$qd = title_q + abstract_q + keywords_q + introduction_q + conclusion_q$$

$$Y_q = 10 \times (current\_year - publication\_year\ of\ q)$$

$W_{qd,T} = LDA(q_d)$
Sort $W_{qd,T}$ *in descending order by weights*
$Y_r = []$
$Zr = []$
for topic t in υ
$F_r [] = E_{r,t} = w_{r,t}$
$Y = 0, q_n = 0$
for each publication q of r if $W_{q,t}$ *is maximum*
$y += y + Yq$
$Y_{r,t} = 1/(10*q_n) * Y$
$Y_r[] = Y_{r,t}$
$Zr,t = \alpha * E_{r,t} + \beta * Y_{r,t} + \gamma * Ar$
$Z_r[] = Z_{r,t}$
$F[] = F_r$
$Y[] = Y_r$
$Z[] = Z_r$
4. *End*

5.6 Ranking

Once manuscripts and reviewers' profiles are built, next step is to rank reviewers based on the proficiency of reviewer for each topic. Profile of manuscript p explored that manuscript p is mainly composition of topics t1, t2, …..tn.
During the reviewer assignment process, the primary constraint is that among assigned reviewers each reviewer ri should have an expertise in topic ti. This leads to more precise reviewer assignment, as manuscripts get reviewed by multiple reviewers who are experts in research fields covered in a manuscript. By keeping same in the mind, Ψ number of most relevant reviewers to each topic in K is obtained. The value of Ψ depends on the number of manuscripts to be reviewed |P| and other constraints like number of reviewers assigned per manuscript (c) and maximum number of manuscripts assigned to each reviewer(m). Here the value of Ψ is set to 25. Based on the expertise values of reviewers to topics (Er, t), from reviewers' profiles top 25 most relevant reviewers selected for each topic, which are further used in final reviewer assignment. Initially, the Brute-Force approach was used for reviewer assignment based on reviewer to topic proficiency value, but it leads to imbalanced reviewer assignment.
Manuscript which is having lower relevance to the topic may get assigned a reviewer which is having higher expertise in the respective topic. To overcome this problem of imbalanced assignment, reviewers are ranked against topics by following the steps described in algorithm 3. Ranking is performed on the basis of metrics defined in reviewers and manuscripts profiles. Initially reviewers are ranked on the basis of expertise level of reviewer to topic (Er, t), which brings most relevant reviewers to topics at the top. Further reviewers are ranked on the basis of proficiency of reviewers for topics.
This is a ranking algorithm we have designed to resolve these issues. Paper which is having lower relevance to the topic may get assigned a reviewer which is having higher expertise in the respective topic. To overcome this problem of imbalanced assignment, reviewers are ranked against topics. Ranking is performed on the basis of reviewers and papers profiles. Reviewers are ranked on the basis of proficiency of reviewers for topics.

5.7 Assignment of Reviewers to Papers

Once the proficiency per topic is computed, and reviewers are ranked, top 5 reviewers as proficient experts are listed per topic. Next important phase is to assign reviewers per paper satisfying the constraints.
The reviewer to manuscript assignment process determines a set of reviewers who are having expertise in topics covered in manuscripts and satisfy the conference specific constraints. To assign a reviewer to a submitted manuscript, the following constraints are satisfied:
1. The reviewer should have expertise in at least one topic covered in a manuscript.

2. The reviewer with a higher proficiency value is preferred.
3. The reviewer should not get assigned a maximum m number of manuscripts.
4. Each manuscript should be reviewed by set of reviewers.
5. Combined expertise of all reviewers cover all paper domains.

Algorithm 4 describes the steps followed for assignment of reviewers to papers.

Algorithm 4: Assignment of Reviewers to Papers
Input: R = {r1, r2, ..., rM}; set of reviewers,
Reviewer profiles with topic-proficiency wise sorted
Output: Reviewer paper assignment { (Pi, Ri1), (Pi, Ri2), (Pi, Ri3), (Pi, Ri4), (Pi, Ri5) }
Wj is workload of jth reviewer , Cov_i is coverage for paper_i

1. *begin*
2. *for each reviewer pi (i=1 to N)*
3. *use Paper Pi profile (Rel_Topic1, Rel_Topic2, Rel_Topic3, Rel_Topic4, Rel_Topic5)*
4. *for k=1 to 5 do*
5. *while j =1 to M do*
6. *For Pi_topick, assign Rj with highest 'proficiency_j'*
7. *Wj=Wj+1*
8. *Pi_Rel_Topick =visited, cov_i=cov_i + 1*
9. *end*

In the earlier step relevant reviewers for each topic obtained, which are ranked on the basis of expertise and proficiency of reviewers for the topic. Ranking is performed in descending order. By keeping in mind the primary constraint as stated in the problem statement section, each reviewer should get assigned a set of reviewers, such that at least one reviewer in a set is having expertise in each key topic covered by manuscript p. Manuscript profile gives the n number of key topics (Γpi) covered and their relevance for manuscript.

Algorithm 4 describes the working procedure of ranking of reviewers based on topics and proficiency. Proficiency is computed using expertise, authority, and topic recency. Ranking uses the matrix defined for profiles of reviewers and papers. Algorithm initially performs ranking in descending order based on proficiency and then assigns these ranked reviewers to topics. Further, the manuscript and reviewer pairs are obtained.

$$A = \{(p1,R1), (p2,R2)....(p|p|,R|P|)\} \quad\ldots\ldots\ldots\ldots(21)$$

6. Results

Experimentation and results showing the performance evaluation along with comparative analysis with state-of-the-art baseline techniques.

6.1 Performance Evaluation Techniques

For measuring the performance, we have selected the performance metrics that are most appropriate and popularly in use. The performance metric that are popularly in use can be broadly classified as- Objective also called as Quantitative measures and Subjective that is also called as Qualitative measures. In a subjective analysis of reviewer assignment systems, after assignments reviewers are requested to provide relevance feedback on the papers they are reviewing as- 'very relevant', 'relevant', 'somewhat relevant' & 'irrelevant'. Sometimes experts are invited to do so. Further the feedback is aggregated to make final decisions on the papers. Whereas In the objective performance measure, the results are compared with expected goal values or the best values named as reference value. In RAP the significant challenge in the RAP is finding the reference value. Often similarity value between paper and expertise of reviewer is used to compute accuracy.

Popular quantitative measures include- efficiency and efficacy.

- Efficiency is computed in terms of Time utilization and Effectiveness is computed as number of Accurate assignments satisfying the constraints.
- Effectiveness is computed by considering the relevance and accuracy.
- Relevance is a measure of the topical similarity between a reviewer candidate and a submission. Most of the researchers have computed the relevance value.
- Accuracy is the degree to which the result of a measurement, calculation, or specification conforms to the correct value or a standard. Here it refers to ratio of total count of accurate assignments to total assignments.

6.2 Performance Metrics

Proposed systems performance is measured using both Subjective and objective metrics. For objective Precision, Recall, F1-Score, Mean Average Precision, Normalized Discounted

Cumulative Gain, and Binary Preference are computed. Generally for accuracy, the empirical evaluations- precision and recall are used and, the performance is compared with manual assignment of reviewers or getting the assignments assessment done by the experts that is subjective analysis.

6.2.1 Precision and Recall

In information retrieval and classification (machine learning), Precision and recall are based on relevance [4, 7]. Precision is the fraction of relevant instances among the retrieved instances. Precision is also known as a positive predictive value. Recall is the fraction of relevant instances that were retrieved. Recall is also known as sensitivity.

Precision takes all retrieved items into account, but it can also be evaluated at a given cut-off rank, considering only the topmost results returned by the system. This measure is called 'precision at k' or 'P@k'. P denotes the number of correctly assigned reviewers and Q denotes the total number of assigned reviewers. N denotes the number of papers.

$$Precision@K = \frac{1}{|P|}\sum_{1}^{|P|}\frac{PA \cap TA}{TA} \quad \dots\dots\dots\dots\dots\dots\dots(22)$$

Where |P| - Number of manuscripts,
PA – Number of accurate assignments (assigned reviewers),
TA - Total number of assignments (assigned reviewers)

Recall- the proportion of the number of relevant reviewer retrieved and Is computed as

$$Recall@K = \frac{1}{|P|}\sum_{1}^{|P|}\frac{PA \cap AA}{PA} \quad \dots\dots\dots\dots\dots\dots\dots\dots(23)$$

Where |P| - Number of manuscripts,
PA – Number of accurate assignments (assigned reviewers),
AA -Number of actual assignments (assigned reviewers)

F1-Score

A measure that combines precision and recall is the harmonic mean of precision and recall [22, 26 30] , the traditional F-measure or balanced F-score. Is computed as

$$\mathrm{F1-score} = 2 * \frac{\mathrm{Precision}*\mathrm{Recall}}{\mathrm{Precision} + \mathrm{Recall}} \quad \dots\dots\dots(24)$$

6.2.2 Mean average precision (MAP)

Mean average precision for a set of papers is the mean of the average precision scores for each paper for reviewer[1, 2, 5]. His computes how much relevant papers are assigned for reviewer and Is computed as

$$\mathrm{MAP} = \frac{1}{|P|}\sum_{1}^{|P|}\frac{1}{\mathrm{Rn}}\sum_{i=1}^{\mathrm{Rn}}\left(\mathrm{Precision@i} * \mathrm{R(ci)}\right)$$

................(25)

Where, |P| is number of manuscripts,
n = φ, number of top reviewers,
Rn = number of eligible reviewers for manuscript p,
R(ci) = 1 if the ith identified reviewers is relevant for the paper p else R(ci)= 0.

6.2.3 Normalized Discounted Cumulative Gain (NDCG)

Normalized discounted cumulative gain uses a graded relevance scale of reviewers from the result set to evaluate the usefulness, or gain, of a document based on its position in the result list and Is computed as [5, 14].

$$NDCG = \frac{1}{|P|}\sum_{1}^{|P|}\frac{\sum_{i=1}^{Rn}\frac{R(ci)}{\log_2(i+1)}}{\sum_{i=1}^{Rn}\frac{1}{\log_2(i+1)}} \dots\dots\dots\dots(26)$$

Where, |P| is number of manuscripts,
n = φ, number of top reviewers,
Rn = number of eligible reviewers for manuscript p,
R(ci) = 1 if the ith identified reviewers is relevant for the paper p else R(ci)= 0.

6.2.4 Binary Preference (bpref)

Bpref measure is a function of how frequently relevant reviewers are retrieved before non-relevant reviewers and is computed as -

$$\mathrm{Bpref} = \frac{1}{|P|}\sum_{1}^{|P|}\frac{1}{\mathrm{Rn}}\sum_{r=1}^{\mathrm{Rn}}\left(1 - \frac{\sum_{i=1}^{r}(1-\mathrm{R(ci)})}{\mathrm{Rn}}\right) \dots(27)$$

Where, |P| is number of manuscripts,
n = φ, number of top reviewers,
Rn = number of eligible reviewers for manuscript p, R(ci) = 1 if the ith retrieved reviewer is relevant to manuscript p else R(ci)= 0.

6.3 Baseline Techniques

There are several baseline methods available to comparative analysis that includes classic algorithms and state-of-the-art techniques[14].

1. ATM- Author Topic Model
2. LM -Language Model
3. LDA LM – Language Model with Latent Dirichlet Allocation
4. TATB – Time Aware and Topic Based Model
5. KCS - Keyword Cosine Similarity
6. BBA - Branch and Bound Algorithm and
7. WMD - Word Mover Distance

6.3.1 Latent Dirichlet Allocation (LDA)

An unsupervised learning technique-LDA considers documents as bags of words neglecting the order of words with assumption that the document is created keeping in mind some set of topics and then a set of words are selected for the specific topic[3]. LDA computes the cosine distance of distribution probability of topics between the reviewer and the manuscript.

6.3.2 Language Model (LM)

The topic of the paper is treated as a query term and the probability of presence of query term with respect to information of experts is computed to find the reviewers that is more appropriate for the paper[3].

6.3.3 Latent Dirichlet Allocation- Language Model (LDA-LM)

In this approach the LDA results and LM are combined to the appropriate reviewers based on the score[2, 3].

6.3.4 Time-Aware and Topic-Based Model (TATB)

Based on LDA, weights are assigned to publications of experts with different weights over time, and further the score is obtained by multiplying it by TF- IDF[3].

6.3.5 KCS

Key-phrase Extraction Algorithm (KEA) is utilized to find the keywords representing the expertise of reviewers and submitted manuscripts and then weights are assigned to the keywords referring to location time[3].

6.3.6 Branch-and-Bound Algorithm (BBA)

This approach uses LDA to obtain the topic distribution of the reviewers and the target papers [3]. The topic distribution of all of the reviewers for a target paper is considered as a whole (a group of reviewers), and the branch-and-bound method is used to quickly determine the appropriate reviewers.

6.3.7 Word2vec-based Word Mover's Distance (WMD) Algorithm

The word embedding of the reviewers and the target papers is computed and using the distance between the text excerpts is computed using EMD (Earth Mover Distance)[3]. It calculates the cosine similarity between the reviewer and the target paper. This value is further used to compute the earth mover distance between the textual data.

6.3.8 Experimentation and Parameter Setting

UPRPAS is a topic modeling and profile based proactive reviewer assignment approach. Topic models for each dataset are built distinctly by keeping LDA hyper parameters constant and varying the number of topics. Constraints are set as the number of reviewers assigned to paper is set to 5 Maximum number of papers assigned to a reviewer is set to 10. Top 5 relevant topics obtained for papers and reviewers' publications. Corpus is formed with tri-gram phrases for building dictionaries using the Bag of Words model. LDA model is built on dictionary by setting hyper-parameters as number of passes are set to 5, chunk size is selected as 100, 500 iterations and 20 passes by Varying K from 5 to 30 as K = 5, 7, 10, 12, 15, 18,20, 25, 30.
Coherence score and perplexity are calculated for each model to find the optimal number of topics. Higher the coherence score means better is the model. Coherence score is obtained for each value of K, for all datasets are shown in table 2.

Table 2: Coherence Score for three Datasets by varying Number of Topics

Dataset	NIPS2019		Interspeech2014		AAAI2014	
Number of Topics (K)	Coherence Score	Perplexity	Coherence Score	Perplexity	Coherence Score	Perplexity
5	0.3119	-8.6089	0.4452	-7.4443	0.4883	-7.6426
7	0.3451	-8.6227	0.4562	-7.4614	0.5142	-7.649
10	0.3671	-7.3443	0.47	-7.4817	0.5093	-7.6629
12	0.3856	-8.1767	0.4599	-7.47	0.499	-7.6596
15	0.4091	-9.6167	0.4323	-10.6798	0.4541	-11.7501
18	0.4131	-18.5601	0.4241	-13.7885	0.4179	-14.1751
20	0.4252	-19.6795	0.4425	-14.5383	0.427	-14.9518
25	0.4304	-11.3786	0.4097	-16.3147	0.4076	16.8542
30	0.4159	-12.7028	0.4318	-18.1167	0.4144	-18.7706

Variation in the coherence score of topic model with respect to number of topics for each dataset can be observed. It is revealed clearly that topic model for datasets NIPS-2019, Interspeech-2014 and AAAI-2014 have highest coherence values for number of topics, K= 25, 10 and 30 respectively. Using topic model, topic distribution for each paper dataset and reviewers' publications are obtained.

Once the topic model is built, papers to topics weight matrix is computed. Table 3 shows papers to topics weight matrix for NIPS dataset for number of topics set to 25.

Table 3: Paper to Topic Weight

Manuscript	Topic1	Topic2	Topic3	….	Topic23	Topic24	Topic25
8296	0.0263	0.0042	0.0063	….	0.0014	0.0069	0.0179
8297	0.1034	0.0047	0.0317	….	0.0015	0.0324	0.0201
8298	0.1168	0.0193	0.0209	….	0.0010	0.0051	0.0295
8299	0.0269	0.0043	0.0065	….	0.0014	0.1641	0.0183
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
8492	0.0555	0.0048	0.0325	….	0.0016	0.0080	0.0459
8493	0.0357	0.0056	0.0086	….	0.0019	0.0392	0.0243

From earlier table this table is constructed showing the Top 5 Relevant Topics for each paper for NIPS Dataset from 25 topics. For example, we can notice that for paper 8492 weight for Topic19 is 0.1836…weight for Topic16 is 0.1357 and similarly weight is 0.0656 for topic 8.

Table 4: Top 5 Relevant Topics to papers for NIPS Dataset

Reviewer	**Rel_ Topic1**	**Rel_ Topic2**	**Rel_ Topic3**	**Rel_ Topic4**	**Rel_ Topic5**
R00001	Topic20 0.4911	Topic25 0.1316	Topic16 0.0888	Topic19 0.0483	Topic18 0.0411
R00002	Topic1 0.1563	Topic16 0.1287	Topic17 0.0839	Topic25 0.0834	Topic20 0.0756
R00003	Topic1 0.1653	Topic6 0.1061	Topic25 0.1019	Topic10 0.0848	Topic12 0.0781
R00004	Topic16 0.0897	Topic17 0.0852	Topic9 0.0831	Topic18 0.0753	Topic5 0.0748
⋮	⋮	⋮	⋮	⋮	⋮
R00103	Topic16 0.1499	Topic19 0.1196	Topic1 0.1092	Topic10 0.09704	Topic25 0.0859
R00104	Topic19 0.1628	Topic1 0.0954	Topic16 0.0843	Topic13 0.0729	Topic25 0.0703
R00105	Topic19 0.1416	Topic1 0.1361	Topic16 0.1161	Topic6 0.1011	Topic8 0.0966
R00106	Topic16 0.1972	Topic19 0.0952	Topic1 0.0854	Topic10 0.0676	Topic17 0.0641

With the help of Paper-Topic weight matrix for top 5 relevant topics, we can find the papers that are most relevant per topic. We can observe topic wise clustering of papers as in pie-chart as shown in figure 6.

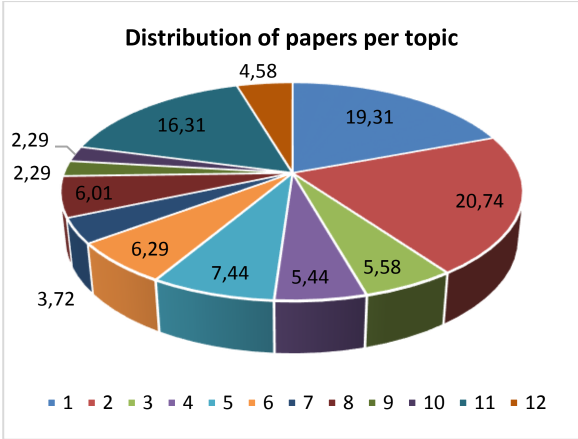

Fig. 6: Distribution of Papers

Figure 7 shows the number of papers relevant to each topic by considering 5 dominant topics.

For example for first relevant topic as topic 16 is for around 126 papers.

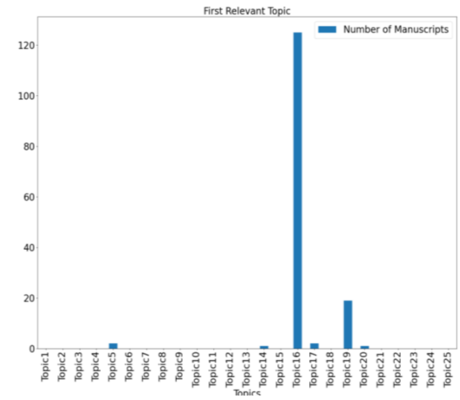

Fig.7: 1^{st} Relevant Topic

Figure 8 shows that topic1 is 4^{th} relevant for 68 count of papers.

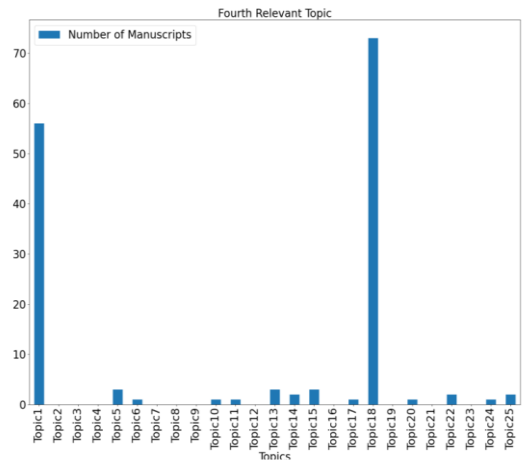

Fig. 8: 4^{th} Relevant Topic

We can visualize topic as word cloud as shown in figure 9.

Fig. 9: Word Cloud Representation for Interspeech2014

In order to evaluate the assignment obtained, for each paper and reviewer pair, similarity scores between paper description and reviewer description are obtained and rated on a scale as shown in table. Each paper reviewer assignment pair along with similarity score and accuracy label is obtained. Once the paper to topic weight is computed, then topic to reviewer publications weight is also computed to build reviewer profiles. Table 5 shows reviewer publications to topic weights for NIPS with Topics =25.

Table 5: Reviewer to Topic Expertise Matrix (F) for NIPS-2019 Dataset @K=25

Reviewer	Topic1	Topic2	Topic3	….	Topic23	Topic24	Topic25
R00001	0.0212	0.0013	0.0086	….	0.0004	0.0021	0.1316
R00002	0.1563	0.0177	0.0376	….	0.0128	0.0066	0.0834
R00003	0.1653	0.0080	0.0107	….	0.0002	0.0008	0.1019
R00004	0.0567	0.0012	0.0518	….	0.0010	0.0627	0.0310
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
R00103	0.1092	0.0214	0.0126	….	0.0006	0.0323	0.0860
R00104	0.0954	0.0165	0.0366	….	0.0160	0.0328	0.0703
R00105	0.1361	0.0113	0.0482		0.0003	0.0120	0.0510
R00106	0.0854	0.0271	0.0237		0.0002	0.0275	0.0595

Once reviewer publications to topic weight is computed, then the Reviewer to Topic Expertise Matrix is built and the top 5 most relevant topics with weight are shown in table 6.

Table 6: Relevant Topics to Manuscripts for NIPS-2019 Dataset @K=25

Reviewer	Rel_Topic1	Rel_Topic2	Rel_Topic3	Rel_Topic4	Rel_Topic5
R00001	Topic20 0.4911	Topic25 0.1316	Topic16 0.0888	Topic19 0.0483	Topic18 0.0411
R00002	Topic1 0.1563	Topic16 0.1287	Topic17 0.0839	Topic25 0.0834	Topic20 0.0756
R00003	Topic1 0.1653	Topic6 0.1061	Topic25 0.1019	Topic10 0.0848	Topic12 0.0781
R00004	Topic16 0.0897	Topic17 0.0852	Topic9 0.0831	Topic18 0.0753	Topic5 0.0748
⋮	⋮	⋮	⋮	⋮	⋮
R00103	Topic16 0.1499	Topic19 0.1196	Topic1 0.1092	Topic10 0.09704	Topic25 0.0859
R00104	Topic19 0.1628	Topic1 0.0954	Topic16 0.0843	Topic13 0.0729	Topic25 0.0703
R00105	Topic19 0.1416	Topic1 0.1361	Topic16 0.1161	Topic6 0.1011	Topic8 0.0966
R00106	Topic16 0.1972	Topic19 0.0952	Topic1 0.0854	Topic10 0.0676	Topic17 0.0641

One of the novelties of research work is we compute more than one expertise domain of each reviewer utilizing the publications of him or her. These publications are spread across a span of years. We also calculate recency as most recent is given preference. Recency of reviewers for each topic is calculated. Table 7 shows the reviewers topics recency matrix for NIPS dataset.

Table 7: Reviewers-Topics Proficiency Matrix (Z) for NIPS2019 Dataset @K=25

	Reviewer1	Reviewer2	Reviewer3	….	Reviewer104	Reviewer105	Reviewer106
Topic 1	0.1643	0.4041	0.4225	….	0.2477	0.3961	0.2427
Topic 2	0.1543	0.4788	0.2038	….	0.2082	0.1987	0.2135
Topic 3	0.1580	0.3188	0.2052	….	0.2183	0.2172	0.2118
Topic 4	0.1537	0.2000	0.1999		0.2000	0.1931	0.2000
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
Topic 22	0.1555	0.2022	0.2043		0.2050	0.2023	0.3642
Topic 23	0.1539	-0.0036	0.1999		0.2080	0.1932	0.2001
Topic 24	0.1548	0.2033	0.2003		0.2164	0.1990	0.2138
Topic 25	0.3995	0.4517	0.4908		0.2351	0.2186	0.2298

Now once expertise, recency and authority is computed now the one value Proficiency is computed with these 3. Table 7 indicates proficiency computations for each reviewer for total 25 extracted topics with topic modeling.

Table 8 shows 12 Topics and most relevant reviewers with all features of reviewer profile. 'E' indicates topic relevance as expertise, 'R' is topic recency, 'A' indicates authority computed using total publications (tp), h-index(h), total citations(c).

Table 8: Topics and reviewer 1 out of top 25 reviewers with all features of reviewer profile

Topic	Reviewer
Topic1	{'reviewer': 'R00097', 'E': 0.36398855, 'R': 0.8, 'rp': 2, 'A': 0.6046790847140093, 'tp': 2, 'h': 1, 'I': 0, 'c': 2}
Topic2	{'reviewer': 'R00104', 'E': 0.23423617, 'R': 0.229, 'rp': 7, 'A': 1.0, 'tp': 20, 'h': 15, 'I': 15, 'c': 1283}
Topic3	{'reviewer': 'R00038', 'E': 0.28674772, 'R': 0.85, 'rp': 2, 'A': 0.9999944405176665, 'tp': 9, 'h': 7, 'I': 5, 'c': 142}
Topic4	{'reviewer': 'R00048', 'E': 0.25450355, 'R': 0.8, 'rp': 1, 'A': 0.9999998860493201, 'tp': 12, 'h': 6, 'I': 4, 'c': 211}
Topic5	{'reviewer': 'R00058', 'E': 0.3240014, 'R': 0.8, 'rp': 2, 'A': 0.9999999342561737, 'tp': 19, 'h': 8, 'I': 7, 'c': 197}
Topic6	{'reviewer': 'R00071', 'E': 0.5573622, 'R': 0.7, 'rp': 3, 'A': 0.9999999999989686, 'tp': 14, 'h': 6, 'I': 5, 'c': 392}
Topic7	{'reviewer': 'R00084', 'E': 0.3375761, 'R': 0.7, 'rp': 2, 'A': 0.9999999999978701, 'tp': 10, 'h': 5, 'I': 4, 'c': 390}
Topic8	{'reviewer': 'R00001', 'E': 0.4908596, 'R': 0.7, 'rp': 1, 'A': 0.9758729785823308, 'tp': 3, 'h': 1, 'I': 1, 'c': 50}
Topic9	{'reviewer': 'R00092', 'E': 0.089189015, 'R': 0, 'rp': 0, 'A': 0.9999774555703496, 'tp': 13, 'h': 5, 'I': 3, 'c': 134}
Topic10	{'reviewer': 'R00031', 'E': 0.30757242, 'R': 0.8344, 'rp': 3, 'A': 0.9536908501514997, 'tp': 6, 'h': 2, 'I': 2, 'c': 30}
Topic11	{'reviewer': 'R00026', 'E': 0.3156765, 'R': 0.825, 'rp': 4, 'A': 0.9999999984538532, 'tp': 14, 'h': 5, 'I': 3, 'c': 287}
Topic12	{'reviewer': 'R00025', 'E': 0.20959224, 'R': 0.6, 'rp': 1, 'A': 0.999907434362447, 'tp': 4, 'h': 4, 'I': 4, 'c': 115}

Table 8 depicts further computations for Relevant top 5 Topics of each Reviewers. It can be noticed that most relevant topics for reviewer with id RP0001 are topics 20,25,16,19,18… Whereas for reviewer id as 103 the relevant topics are topic 16, 19, 1, 10, 25.

Table 9: Relevant Topics to Manuscripts for NIPS-2019 Dataset @K=25

Reviewer	Rel_Topic 1	Rel_Topic2	Rel_Topic3	Rel_Topic 4	Rel_Topic 5
R00001	Topic20 0.4911	Topic25 0.1316	Topic16 0.0888	Topic19 0.0483	Topic18 0.0411
R00002	Topic1 0.1563	Topic16 0.1287	Topic17 0.0839	Topic25 0.0834	Topic20 0.0756
R00003	Topic1 0.1653	Topic6 0.1061	Topic25 0.1019	Topic10 0.0848	Topic12 0.0781
R00004	Topic16 0.0897	Topic17 0.0852	Topic9 0.0831	Topic18 0.0753	Topic5 0.0748
⋮	⋮	⋮	⋮	⋮	⋮
R00103	Topic16 0.1499	Topic19 0.1196	Topic1 0.1092	Topic10 0.09704	Topic25 0.0859
R00104	Topic19 0.1628	Topic1 0.0954	Topic16 0.0843	Topic13 0.0729	Topic25 0.0703
R00105	Topic19 0.1416	Topic1 0.1361	Topic16 0.1161	Topic6 0.1011	Topic8 0.0966
R00106	Topic16 0.1972	Topic19 0.0952	Topic1 0.0854	Topic10 0.0676	Topic17 0.0641

Table 9 indicates top 5 most Relevant Reviewers to Topics. It indicates that for the topic 22 the top 5 most relevant reviewers with respect to their expertise are reviewer id 71, 82, 5,73, and 95.

Table 10: Relevant Reviewers to Topics for NIPS-2019 Dataset @K=25

	Reviewer1	Reviewer2	Reviewer3	….	Reviewer29	Reviewer30
Topic1	R00011 0.2275	R00078 0.1884	R00040 0.1792	….	R00029 0.0976	R00060 0.0965
Topic2	R00054 0.0391	R00033 0.0348	R00088 0.0341	….	R00096 0.0108	R00016 0.0104
Topic3	R00049 0.1126	R00020 0.1041	R00084 0.0721	….	R00014 0.0261	R00024 0.0259
Topic4	R00052 0.0003	R00020 0.0003	R00057 0.0003		R00086 0.0001	R00075 0.0001
:	:	:	:	:	:	:
Topic22	R00071 0.1845	R00082 0.1561	R00005 0.1498	….	R00073 0.0371	R00095 0.0367
Topic23	R00071 0.0329	R00046 0.0316	R00013 0.0256	….	R00040 0.0018	R00093 0.0017
Topic24	R00088 0.0889	R00028 0.0697	R00004 0.0627	….	R00081 0.0255	R00016 0.0254
Topic25	R00066 0.1390	R00050 0.1375	R00001 0.1316		R00104 0.0703	R00082 0.0685

Paper to reviewer assignment is done by first ranking the reviewers based on their proficiency value. Along with proficiency value, care has been taken that all constraints are met. The value of the minimum number of reviewers to be assigned to each paper is set to 5 and no reviewer is assigned more than 10 papers. Table 11 shows the assignment of the most appropriate 5 reviewers assignment for each paper. While assigning papers the constraints are satisfied such that almost each topic of papers is covered, no reviewer is assigned more than 10 papers and each paper is assigned 5 reviewers.

Table 11: Manuscript Reviewer Assignment

Assigned Reviewers / Manuscript	Reviewer1	Reviewer2	Reviewer3	Reviewer4	Reviewer5
8296	R00048	R00029	R00019	R00035	R00096
8297	R00081	R00106	R00025	R00059	R00047
8298	R00095	R00100	R00043	R00055	R00076
8299	R00095	R00004	R00101	R00040	R00056
:	:	:	:	:	:
8492	R00034	R00065	R00015	R00072	R00085
8493	R00092	R00016	R00093	R00076	R00062
8494	R00008	R00029	R00019	R00027	R00083
8495	R00061	R00029	R00055	R00014	R00104

Accuracy Labels Generation

In order to evaluate the assignments obtained from the reviewer assignment system ground truth is required. It is a very brainstorming, complex, time-consuming and error prone task to generate ground truth labels manually. To overcome this, an automatic evaluation approach is proposed.

This is again our novel contribution. The labels for each paper and reviewers are computed that are also called as paper descriptions and reviewer descriptions.
Lp and Lr are the obtained by taking union of set of top 20 labels of 5 the most relevant topics for paper (p) and reviewer (r).

CS=|{ Lp }∩{Lr} |/|{ Lp } ∪ {Lr}| ……….(23)

Union of set of top 20 labels of 5 the most relevant topics for paper and reviewer is computed to represent a similarity score. and then the accuracy labels are assigned.
The label as Very Relevant if similarity score is higher than 85 %, labeled as relevant when similarity score is between 65 to 85%. When the similarity score is between 50 to 65% then label Somewhat Relevant is assigned. And label is Irrelevant if score is below 50%.

- The accuracy labels are assigned as- (*cs* is similarity score)
  - Very Relevant (V) for CS>85 %
  - Relevant (R) for 65%<=CS<85%
  - Somewhat Relevant (SR) for 50%<=CS<65%
  - Irrelevant (I) for CS<50 %

Table 12 shows paper's Reviewers' Assignments with Similarity Scores and Accuracy Labels for NIPS.

Table 12: Manuscript's Reviewers' Assignments with Similarity Scores and Accuracy Labels

Assigned Reviewers / Manuscript	Reviewer1	Reviewer2	Reviewer3	Reviewer 4	Reviewer 5
8296	R00048 0.7075 Relevant	R00029 0.7389 Relevant	R00019 0.7195 Relevant	R00035 0.7489 Relevant	R00096 0.6669 Relevant
8297	R00081 0.8943 Very Relevant	R00106 0.8277 Very Relevant	R00025 0.8615 Very Relevant	R00059 0.7532 Relevant	R00047 0.7842 Relevant
8298	R00095 0.5813 Some What Relevant	R00100 0.6498 Relevant	R00043 0.5558 Some What Relevant	R00055 0.5123 Some What Relevant	R00076 0.7856 Relevant
8299	R00048 0.7075 Relevant	R00004 0.5351 Some What Relevant	R00101 0.7083 Relevant	R00040 0.5818 Some What Relevant	R00056 0.7597 Relevant
8492	R00034 0.7507 Relevant	R00065 0.7844 Relevant	R00015 0.8885 Very Relevant	R00072 0.8192 Very Relevant	R00085 1 Very Relevant
8493	R00092 0.8969 Very Relevant	R00016 0.8072 Very Relevant	R00093 0.8485 Very Relevant	R00076 0.7798 Relevant	R00062 0.9021 Very Relevant
8494	R00008 0.7124 Relevant	R00029 0.7859 Relevant	R00019 0.8339 Very Relevant	R00027 0.909 Very Relevant	R00083 0.9075 Very Relevant
8495	R00061 0.5461 Some What Relevant	R00029 0.3905 Irrelevant	R00055 0.3148 Irrelevant	R00014 0.5363 Some What Relevant	R00096 0.6669 Relevant

Figure 10 shows graphical representation of first five reviewers and Accuracy Labels for NIPS. It can be noticed that count of very relevant + relevant + somewhat relevant labels is more than 85%.

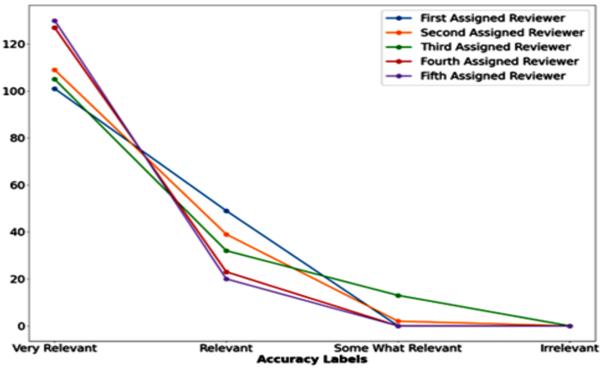

Fig.10: Assignment Relevancy Labels and Reviewer-Manuscript Assignments for NIPS-2019 Dataset @K=7 and @F=30

Figure 11 is the same graph for interspeech dataset. It indicates that though the very relevant labels for 4th and 5th reviewer are comparatively lesser, still the overall labels are very relevant and relevant . This is the same graph for the AAAI dataset.

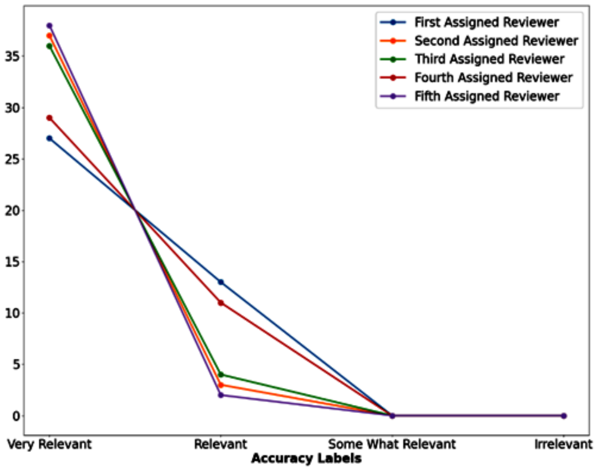

Fig. 11: Assignment Relevancy Labels and Reviewer-Manuscript Assignments for Interspeech2014 Dataset @K=7 and @F=30

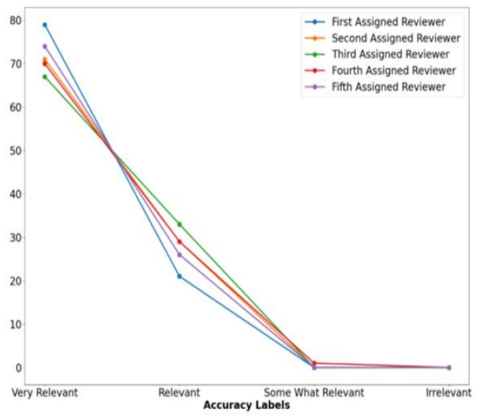

Fig. 12: Assignment Relevancy Labels and Reviewer-Manuscript Assignments for AAAI2020 Dataset @K=7 and @F=30

- Most of the researchers have measured the accuracy in terms of average similarity for reviewers assignment.
- Figure is boxplot for Average Similarity Values for UPRPAS & DPRPAS for 3 datasets NIPS, Interspeech and AAAI.
- From figure 13, it can be noticed that very few outliers and we can notice even distribution.

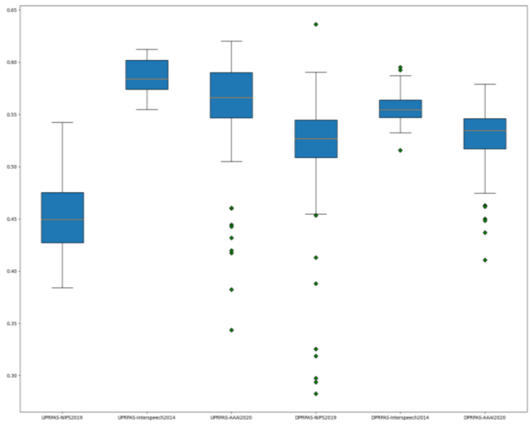

Fig.13: Average Similarity Values for UPRPAS & DPRPAS NIPS, Interspeech and AAAI

In addition to performance metrics, to confirm the coverage of topic, and multi-domain analysis, topic coverage is calculated. Topic coverage for reviewer paper assignment signifies that reviewers assigned to paper satisfy that at least one reviewer having an expertise in at least one topic among all topics covered in the papers. The average topic coverage obtained for all three datasets are shown in table.

rom table 13, it is observed that the topic coverage ranges between 81% to 100%. For all three datasets, average topic coverage is obtained and is above 80% , that means paper reviewer assignments satisfy the multi-domain analysis and constraint topic coverage leading to more accurate assignments and reviews.

Table 13: Average Topic Coverage

Dataset	NIPS	Interspeech	AAAI
No. of papers	150	40	199
No. of Reviewers	106	106	106
UPRPAS - Avg. Topic Coverage	85.87%	98%	100%

The performance of proposed work with existing baseline techniques is compared as shown in table 14, 15 and 16.

Table 14: Performance Metrics Comparison with Baseline Algorithms for NIPS2019

Method	Precision	Recall	F1-score	MAP	NDCG	BPREF
LDA	0.3153	0.2331	0.2681	0.1847	0.3654	0.6695
LM	0.4433	0.3297	0.3782	0.3475	0.5282	0.7603
LDA-LM	0.4446	0.3306	0.3791	0.3484	0.5292	0.7607
TATB	0.3306	0.2445	0.2811	0.2089	0.3902	0.6838
KCS	0.0293	0.0218	0.0250	0.0082	0.0340	0.4919
BBA	0.0389	0.0286	0.0329	0.0268	0.0810	0.5072
WMD	0.4482	0.3327	0.3807	0.3809	0.5320	0.7647
SPM-RA	0.6319	0.4703	0.5393	0.5784	0.7198	0.8773
UPRPAS	0.9400	0.4908	0.6449	0.7951	0.8855	0.9377

Table 15: Performance Measure Comparison with Baseline Algorithms for Inter-speech2014 Dataset @K=10 and @F=30

Method	Precision	Recall	F1-score	MAP	NDCG	BPREF
LDA	0.3330	0.2005	0.2503	0.2036	0.3417	0.6555
LM	0.4680	0.2828	0.3526	0.3281	0.5027	0.7422
LDA-LM	0.4670	0.2823	0.3519	0.3271	0.5019	0.7419
TATB	0.3365	0.2029	0.2531	0.2103	0.3492	0.6578
KCS	0.1325	0.0836	0.1025	0.0611	0.1449	0.5522
BBA	0.1245	0.0715	0.0908	0.0497	0.1502	0.5535
WMD	0.4570	0.2733	0.3420	0.3101	0.4897	0.7319
SPM-RA	0.5435	0.3275	0.4087	0.4137	0.5790	0.7833
UPRPAS	**0.824**	**0.4716**	**0.5999**	**0.8184**	**0.8517**	**0.9334**

Table 16: Performance measure comparison with existing systems for AAAI-2014 dataset @K=7 and @F=30

Method	Precision	Recall	F1-score	MAP	NDCG	BPREF
LDA	0.3330	0.2005	0.2503	0.2036	0.3417	0.6555
LM	0.4680	0.2828	0.3526	0.3281	0.5027	0.7422
LDA-LM	0.4670	0.2823	0.3519	0.3271	0.5019	0.7419
TATB	0.3365	0.2029	0.2531	0.2103	0.3492	0.6578
KCS	0.1325	0.0836	0.1025	0.0611	0.1449	0.5522
BBA	0.1245	0.0715	0.0908	0.0497	0.1502	0.5535
WMD	0.4570	0.2733	0.3420	0.3101	0.4897	0.7319
SPM-RA	0.5435	0.3275	0.4087	0.4137	0.5790	0.7833
UPRPAS	0.9840	0.9900	0.9870	1.0000	1.0000	1.0000

Comparative analysis of performance of existing state-of the art systems and proposed systems is presented at table 17. It reveals that the proposed system exhibits the improved performance even for AAAI 2014 dataset. The table 17 presents the performance for the proposed system for the three datasets.

Table 17: Performance measure for NIPS, Interspeech ad AAAI-2014 dataset @K=7 and @F=30

Dataset	Precision	Recall	F1-score	MAP	NDCG	BPREF
NIPS 2019	0.824	0.4716	0.5999	0.8184	0.8517	0.9334
Interspeech 2014	0.9400	0.4908	0.6449	0.7951	0.8855	0.9377
AAAI 2014	0.9840	0.9900	0.9870	1.0000	1.0000	1.0000

7. Conclusions

The proactive novel system is designed as a solution for the Reviewer Assignment and has presented a novel algorithm and term proficiency that helps to identify the expert as the weighted average of authority, expertise, and recency as the most accurate reviewer to the paper and it reduces computational complexity. Experimental results and performance analysis reveal that the average cumulative similarity also known as an affinity that measures relevance between a paper and the reviewers is improved and ranges between 0.78 to 0.99. The results indicate that the proposed UPRPAS demonstrates a higher accuracy as compared to baseline techniques. Accuracy for NIPS is 77.60%, for AAAI is 74.90% and for Interspeech 4 is 83.50%. The results indicate that the proposed DPRPAS demonstrates a higher accuracy as compared to baseline techniques. Accuracy for NIPS is 74.33%,for AAAI is 65.75%. and for Interspeech is 68.75%. The assignment assures the satisfaction of constraints- load, coverage, and topic coverage. The recent practice of blind reviews assures that there is no conflict of interest. The topic coverage is achieved from 80% to 100%.

Along with reviewer assignment system for articles, journal and conference papers, the proposed system has applicability in a wide set of applications that include- Patient doctor assignment, Matching funding agencies to research proposals, Assigning managers to construction projects, Course-teacher assignment, Conference or Journal papers topical analysis, Candidates to Interviewer assignment and many similar.

Key Research Contributions

The research work aims at providing a novel and efficient solution for the reviewer assignment problem. Key contributions are listed below-

i) A Proactive Reviewer Paper Assignment System (PRPAS) is developed as a solution to the reviewer assignment problem and is made available as a product for others to test and use. https://apras.herokuapp.com for use by researchers, journal editors, and the conference chairs. Further the feedback and suggestions can be positively used for improvement in the system.

ii) The Test dataset and ground-truth datasets built are made available at Kaggle for researchers to use it at https://www.kaggle.com/abolihpatil/dataset-reviewer-paper-assignment-problem-ahp-pnm.

8. Future Scope

The text of the paper can be utilized to build a topic model instead of text from a few sections leading to higher accuracy. Dictionary for research domains can be prepared as a master gold standard list of topics to compare against every corpus as a ready reference for researchers. A master database of reviewers with topic labels that are regularly updated can be built and made available.

References:

[1] Camillo J. Taylor, On the Optimal Assignment of Conference Papers to Reviewers, *IEEE Conference on Computer Vision and Pattern Recognition*, 2008, pp.1-6.

[2] Cheng Long, Raymond Chi-Wing Wong, Yu Peng, Liangliang Ye, On Good and Fair Paper-Reviewer Assignment, *IEEE International Conference on Data Mining, The Hong Kong University of Science and Technology*, 2013, pp.1145-1150.

[3] David Mimno, Andrew McCallum, Expertise Modeling for Matching Papers with Reviewers, ACM, KDD, *North East Student Colloquium on Artifcial Intelligence (NESCAI),* 2007, pp.1-10.

[4] David B. Resnik, Susan A. Elmore, Conflict of Interest in Journal Peer Review, *SAGE Journal Taxicologic Pathology*, Vol. 46(2), (2018), pp.112-114.

[5] Dinh Tuyen Hoang, Ngoc Thanh Nguyen, Dosam Hwang, Decision Support System for Assignment of Conference Papers to Reviewers, *International Conference on Computer Communication and Informatics ICCCCI* ,2019, pp. 441–450.

[6] Don Conryy, Yehuda Korens, and Naren Ramakrishnan, Recommender Systems for the Conference Paper Assignment Problem, *The ACM Recommender Systems conference (RESSYS)*, 2009, pp. 1-8.

[7] Emaad Manzoor, Nihar B. Shah, Uncovering Latent Biases in Text:Method and Application to Peer Review, *The Thirty-Fifth Conference on Artificial Intelligence(AAAI)*, 2021, pp. 1-18.

[8] Fan Wang1, Sun Yat-Sen, A Survey on Reviewer Assignment Problem, *The 31st International Conference on Industrial, Engineering and Other Applications of Applied Intelligent Systems(IEA/ AIE),* 2008, pp. 718-727.

[9] Federico Bianchi,Flaminio Squazzoni, Is there Bettter Than One: Simulating the Effect of Reviewer Selection and Behavior on the Quality and Efficieny of Peer Review, *Proceedings of the 2015 Winter Simulation Conference*, 2015, pp.4081-4089

[10] Graham Neubig,Pengfei Liu, Weizhe Yuan CanWe Automate Scientific Reviewing, *arxiv, 2021,* pp.1-29.

[11] Hongwei Peng, Haojie Hu, Keqiang Wang, and Xiaoling Wang, Time-Aware and Topic-Based Reviewer Assignment, *22nd International Conference on Database Systems for Advanced Applications(DASFAA),* 2017, pp. 145–157.

[12] Ivan Stelmakh, Nihar B. Shah and Aarti Singh, Peer Review 4 All: Fair and Accurate Reviewer Assignment in Peer Review, *Proceedings of the 30th International Conference on Algorithmic Learning Theory*, 2018, pp.1-47.

[13] Shicheng Tan, Zhen Duan, Shu Zhao, Chen· Yanping Zhang, Improved reviewer assignment based on both word and semantic features, *Information Retrieval Journal*, 2021, Volume 24, pp.175–204

[14] Kevin Leyton-Brown, Mausam, Yatin Nandwani, Hedayat Zarkoob, Chris Cameron, Neil Newman, Dinesh Raghu, Matching papers and reviewers at large conferences, *Journal of Artificial Intelligence,*Volume 331, 2024, 104119,pp. 1-24.

[15] Ruihan Wu, Chuan Guo, Felix Wu, Rahul Kida mbi, Laurens VanDeraten, Kilian Weinberger, Making paper reviewing robust to bid manipulation attacks', *International Conference on Machine Learning, PMLR*, 2021, pp. 11240-11250