CS771A - Project Proposal
Project Title - Reviewer Recommender
Group - 48 - MLTeens
Mentor - Dr. Purushottam Kar

Amrit Singhal (amrits@iitk.ac.in)
Anmol Porwal (anmolp@iitk.ac.in)
Harshvardhan (harshv@iitk.ac.in)
Homanga Bharadwaj (homangab@iitk.ac.in)
Varun Khare (varun@iitk.ac.in)

***Abstract*— The currently implemented system for reviewing an appropriate recommender for any submitted paper at a conference is most commonly based on the Toronto Paper Matching System, that uses standard LDA based modelling. Practically, the recommendations made through it fail to include the most relevant reviewers who work in areas closely related to that of the paper. We aim to develop a better generative model using modified LDA that does not concentrate its attention on words, but on sentences as the primitive units. As LDA is a count based-algorithm, we plan on implementing sentence vectors to get semantically close sentences in one 'class'. This allows a reliable extraction of the contexts of the paper and consequently allowing us to perform a better comparison between the paper and previously existing papers, obtaining a more realistic score for the sematic proximtiy of papers and authors. It also enables extraction of the specific closely related works mentioned in the paper, which should be given a higher weight for a potential reviewer.**

I. PROBLEM STATEMENT

The problem statement for this project is, given a paper and a list of available reviewers, recommend the reviewers best suited for reviewing the given paper. More formally given a set of N reviewers, we need to find the best k reviewers for a given fresh paper using the set of resources mentioned in the Resources section. We would currently be working in a scenario where all papers will be same or close to the NIPS format and each of these papers have an explicit Experiment Section or a Related works section, although our approach would be general in nature, which can be applied to any paper with slight modifications. We would be doing the generalised version depending on time constraints.

II. EXISTING WORK

The most popular and prevalent technique employed by most of the top conferences like NIPS, ICML CVPR for automated paper-reviewer assignment is the Toronto Paper Matching System. There have approaches based on using graphs (eg citation graphs) to better represent the context of the paper, however, all these approaches tend to use standard LDA for generative topic modelling and hence fail to utilize the information available properly. The one common drawback for the existing paper-matching systems is a very high-level topic modeling of the paper. They tend to generate keywords for the paper and match based on these when instead, they should be comparing the experiments and related works to get a more fine tuned recommendation.

A. Toronto Paper Matching System

TPMS generates a score for the quantifying the matching for a given paper-reviewer pair. This score is first generated based on the reviewer's published papers and the submission in an unsupervised fashion. After the initial scores are generated, further tuning of the scores is done by guiding a score elicitation procedure (e.g., using active learning to query scores from reviewers).Using these scores we complete the reviewer-submission score matrix and assign appropriate reviewers according to other constraints.

For the unsupervised part, the score is calculated by taking the dot product of a feature representation of the reviewer's existing work and the submitted paper. This feature representation is obtained using a standard bag-of-words frequency counts or by employing Latent Dirichlet Allocation for topic modeling.

For the supervised part, a few approaches have been described. The simplest one is using linear regression where a separate model is trained for each reviewer. This approach fails in the case of conferences because a lot of reviewers are likely to have very few or zero elicited scores. To account for this, a system of parameter sharing is proposed wherein global, reviewer-specific and paper-specific parameters are simultaneously utilized. Another way to view this is as a collaborative filtering task wherein Probabilistic Matrix Factorization is utilized. This PMF approach performs poorly in the cold-start regime.

The drawbacks of this System are primarily because of how it models expertise scores and its matching procedure(taking dot product). Topic modeling approaches of bag-of-words frequency counts and LDA do not take into account the context deeply enough. The models they

produce are based on glorified "counting" procedures and neither do they take the semantic aspect of the paper nor the specificities of the paper-reviewer matching problem into account. Thus these fail to exploit certain important details like the mentioning of related works or the comparisons of experimental results. The core topic or problem which the paper addresses might not make it as high up as the general theme of the paper. The similarity scores are taken by a simple dot product of frequency counts which may give high scores to reviewers of related fields to the general theme of the paper but not at all related to the specific problem the paper wishes to address. Similarly, the coauthors of references and those mentioned as related works may get a lower rank.

B. Robust Paper-Reviewer Assignment Model [3]

This is yet again another approach towards evaluating the similarity scores to be used for ranking the reviewers for each paper. The paper constructs a graph of the paper to be reviewed along with all the reviewers as the nodes. During evaluation of the similarity score the reviewers expertise, authority and his background knowledge is taken into account.

For assigning topic weights different modelling is done to assign reviewer-reviewer connection (number of common co-authors), and reviewer-query connection (LDA topic modelling). They used Random Walk with Restart (RWR using sparsity constraint) to assign the weights in the graph based on the similarity score generated earlier.

The major drawback in this approach is again using LDA topic modeling for score evaluation with the query paper. Though they gave a novel approach to utilising the expertise and authority of a reviewer, their model again failed at comparing the reviewers with the query paper. At he same time this paper didn't utilize the citations already present in the paper for score evaluation. These citations come handy in identifying the related works of the paper which can then be used in assigning the edge weights. Just like them, we too feel the power of graphs can be leveraged to a much higher level in representing the context of the paper and its association with other papers/reviewers and hence we seek to implement citation graphs for each paper and then convolve with the reviewers' graph to generate the score for ranking. We would also be using modified LDA (sentence2vec) for topic modeling and hence intuitively should gain an upper-hand in the prediction of ideal reviewer.

We thought of using sequential LDA too for topic modelling but the effect of order in the words seems much less when we already define the constituent words. The rules of the language (grammar) primarily define the order of the words and hence superficially it does not seem to be of much importance in defining the context of the sentence.

III. Resources

We will be making use of the datasets used in the TPMS paper as this would allow us to compare our model with the SOTA systems currently available. Additionally we will also procure paper-reviewer pairs and most importantly feedback of the reviews from the NIPS website. The feedback and the reviewer rating will be used for training the recommendation system. We reason that these metrics represent the most suited reviewer for the paper and hence will serve as a reliable source for validating out model. We will also be using the recommended reviewers section given by the authors of the paper to be reviewed along with the bidding done by the reviewers[8] in order to achieve a more realistic evaluation of the predicted reviewers.

IV. Proposed Approach

We try to include some of the more intuitive biases about the paper-reviewer matching task and about a typical research paper in general. Generally, a paper will be competing against a maximum of 4-5 papers and/or methods. Identifying these papers or methods and their respective authors or experts, simplifies the reviewer assignment problem to a large extent since these individuals can serve as the best reviewers for the given paper. A more careful analysis of the related works section or the experiment section of a paper can provide very important clues to finding these experts. The reason why other existing methods cannot utilize this inherent information present in the paper's content is because of too high-level topic modeling. We try to solve this problem by tweaking how the topic models are generated. We propose a Sentence2vec model which can take into account context and not just counts. Also by considering whole sentences instead of words we will be better able to gauge contexts taking verbs, subjects and objects equal importance as opposed to LDA where the subjects and objects get higher preferences.

We will extract the related works and experiments section from the paper and give higher weights to words and citations appearing in the related works section and the experiments section. This increases the contributions of these sections in the final topic models/citation graphs. After this we use a citation-graphs approach so that the cited co-authors get preference in the reviewer selection process.

For incorporating citation graphs we can use the following implementation schemes The first approach is will also incorporate some parts of the Robust Paper Reviewer assignment model[3] for the citation graphs with some changes. This will include the tweaked topic models generated from the Sentence2vec approach. The weights between nodes can be calculated as the similarity score between the papers. The citation graph for each reviewer is the complete graph of all his publications with the weights between nodes calculated as earlier.
For any paper the citations can be classed into a few major groups having higher weighted edges between their nodes

while edges between nodes of different groups would be having lower weights. The groups will generally have a common topic label which can be found out easily from the topics of group's constituent papers.

For each paper-reviewer pair we will use their similarity score between their respective graphs. We will rank all the authors in our reviewer list by matching their graphs with the paper graphs with highest rank given to authors with highest similarity scores.
The inclusion of weights in the citation graphs and the topics decreases the likelihood of important cited reviewers landing up low on the review list.
Possible extensions to this proposed work include

- This method can only cover those authors who have been cited in the paper we are concerned with and might miss authors who are relevant but haven't been cited. To account for this, we can enforce a rule which is, search along high weighted edges. If we get an author with high similarity to some group then we try to search in the group of that author for more relevant citations or authors,i.e, if an author has similarity score K then each author of every paper in that group of the author will receive an increase in their scores proportional to their distance from the relevant paper and K.

- Sentence2vec approach can further be extended to identifying experimental or related works sections and even instances where comparisons with other papers are made in the query paper by training a classifier to do just this.

REFERENCES

[1] Laurent Charlin, Richard Zemel, The Toronto Paper Matching System: An automated paper-reviewer assignment system, in: ICML Workshop on Peer Reviewing and Publishing Models (PEER), 2013.
[2] David M. Blei, Andrew Y. Ng, and Michael I. Jordan. 2003. Latent dirichlet allocation. J. Mach. Learn. Res. 3 (March 2003), 993-1022.
[3] Xiang Liu, Torsten Suel, and Nasir Memon. 2014. A robust model for paper reviewer assignment. In Proceedings of the 8th ACM Conference on Recommender systems (RecSys '14). ACM, New York, NY, USA, 25-32. DOI:https://doi.org/10.1145/2645710.2645749.
[4] Chen, Edwin. Introduction to Latent Dirichlet Allocation. blog.echen.me/2011/08/22/introduction-to-latent-dirichlet-allocation/.
[5] Salakhutdinov, Ruslan and Mnih, Andriy. Probabilis- tic matrix factorization. In Platt, J.C., Koller, D., Singer, Y., and Roweis, S. (eds.), Advances in Neu- ral Information Processing Systems 20 (NIPS) , pp. 12571264. MIT Press, Cambridge, MA, 2008.
[6] Rigaux, Philippe. An iterative rating method: applica- tion to web-based conference management. In SAC , pp. 16821687, 2004.
[7] Conry, Don, Koren, Yehuda, and Ramakrishnan, Naren. Recommender systems for the conference paper assignment problem. In Proceedings of the Third ACM Conference on Recommender Systems (RecSys-09) , pp. 357360, New York, New York, USA, 2009. ACM. ISBN 978-1-60558-435-5.
[8] https://cmt.research.microsoft.com/cmt/
[9] https://nips.cc/Conferences/2013/PaperInformation/ReviewerInstructions
[10] arXiv:1407.7073 [cs.GT]
[11] Taylor, Camillo J. On the optimal assignment of con- ference papers to reviewers. Technical Report MS- CIS-08-30, UPenn, 2008.