CSRA: The C³SR Reviewer Assignment System

Omer Anjum,[1] Hongyu Gong,[1] Qiang Ning,[1,3] Suma Bhat,[1] Jinjun Xiong,[2] and Wen-Mei Hwu[1]

[1]University of Illinois at Urbana-Champaign, [2]IBM T. J. Watson Research Center, [3]Allen Institute for Artificial Intelligence

Motivation

Peer review is crucial for scientific achievement. The conventional reviewer assignment process for academic conferences is now challenged by the fast-growing number of paper submissions (Fig 1): while timely review for all submissions is necessary, manual reviewer assignment is extremely labor-intensive due to the following reasons.

- **Time-consuming**: To judge the expertise of a reviewer on a specific submission, one often needs to learn the reviewer's research from his/her publication records.
- **Not scalable**: #judgments = #reviewers x #submissions
- **Optimization difficulty**: The assigning process is often an optimization problem with thousands of variables which also considers reviewer's load and review quality.
- **Limited diversity:** For a small reviewer pool, limited support is available to cover the diversity of topics in new submissions which is growing rapidly.

Way Forward

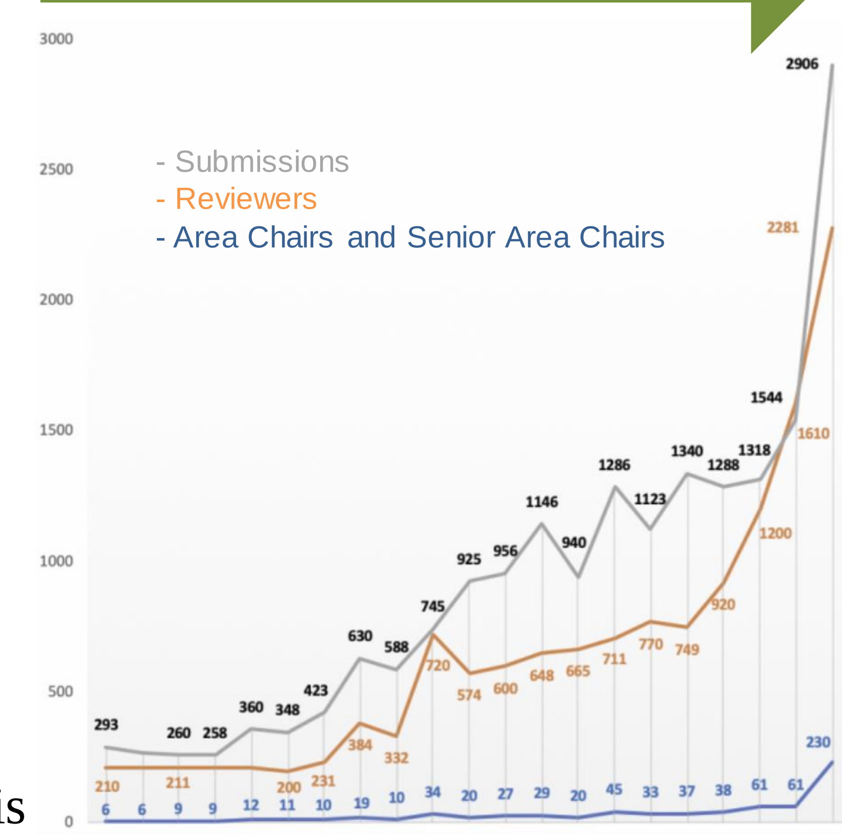


Figure 1: Rapid growth of the participation in the annual meeting of the Association for Computational Linguistics ACL [1]

Automated Reviewer Assignment

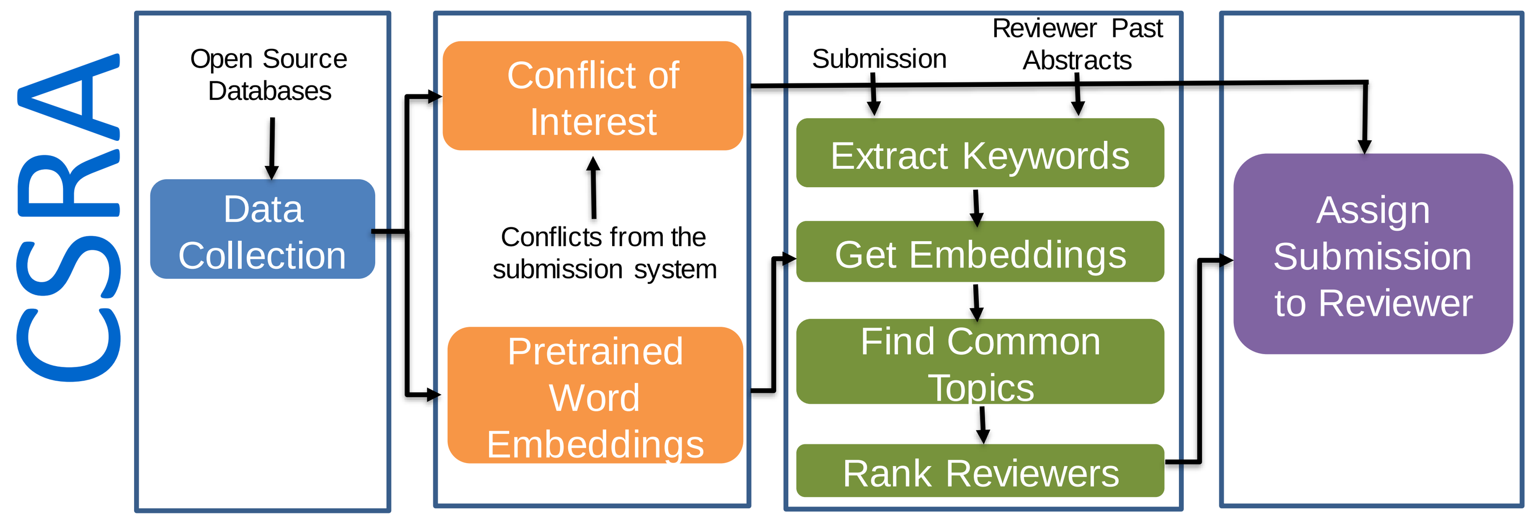
To address those issues, we propose to automate the process via modern *natural language processing* (NLP) and *optimization* techniques. Specifically, we expect to automatically

- judge the expertise level of each reviewer to each submission, by embedding all the submissions and the reviewers within the same semantic space
- assign a required number of expert reviewers to each submission by formulating the assignment process as an optimization problem, which can also
  - avoid conflicts of interests
  - balance the load of different reviewers

Existing Solutions

- LDA based models
  - Mechanism: probability distribution over topics
  - State-of-the-art examples: [2,3], TPMS[4],
  - Limitation: fail to capture lexical level similarity
- LSI based models
  - Mechanism: identifying occurrence patterns of words in documents
  - Limitation: rely on lexical overlap to measure document similarity
- TFIDF based models
  - Mechanism: words weighting based on their importance
  - Limitation: constrained to word-level comparison in similarity evaluation

Proposed



Features

- Do not rely on Bag-of-Words models
- Automatic keyword extraction
- Prioritize recent research interests
- No human in the loop
- Easy to use: simply upload submission abstracts and meta-data in the format used by popular systems like "hotcrp"

Common Topic Model

- Reviewer profile: $R = [r_1, r_2, \dots, r_n]$, where r_i is the embedding of the i^th word in their profile
- Reviewer topic vectors: $P = [p_1, p_2, \dots]$
- Reviewer topics are extracted from the profile:
  $P = Ra$, (a is coefficient)
- Submission S : $S = [s_1, s_2, \dots, s_m]$, where s_j is the embedding of the j^th word in a submission
- Submission topic vectors: $Q = [q_1, q_2, \dots]$
- Submission topics are extracted from the submission:
  $Q = Sb$, (b is coefficient)

Reviewer Matching

- Common topic pairs (P, Q) between reviewer R and submission S

$$\max \mathrm{sim}(P, Q)$$
$$\text{s.t.} \quad P = Ra, \quad Q = Sb, \quad P^T P = Q^T Q \text{ (no duplicate topics)}$$

- Reviewer-Topic relevance: $\mathrm{rel}(R, P) = \mathrm{sim}^2(R, P)$
- Submission-Topic relevance: $\mathrm{rel}(S, P) = \mathrm{sim}^2(S, P)$
- Reviewer-Submission relevance:

$$\mathrm{rel}(R, S) = \frac{2 \cdot \mathrm{rel}(R, P) \cdot \mathrm{rel}(S, P)}{\mathrm{rel}(R, P) + \mathrm{rel}(S, P)}$$

Evaluation Metric

- Precision @ Top N or P@N
  - true positive / (true positive + false positive)

Data

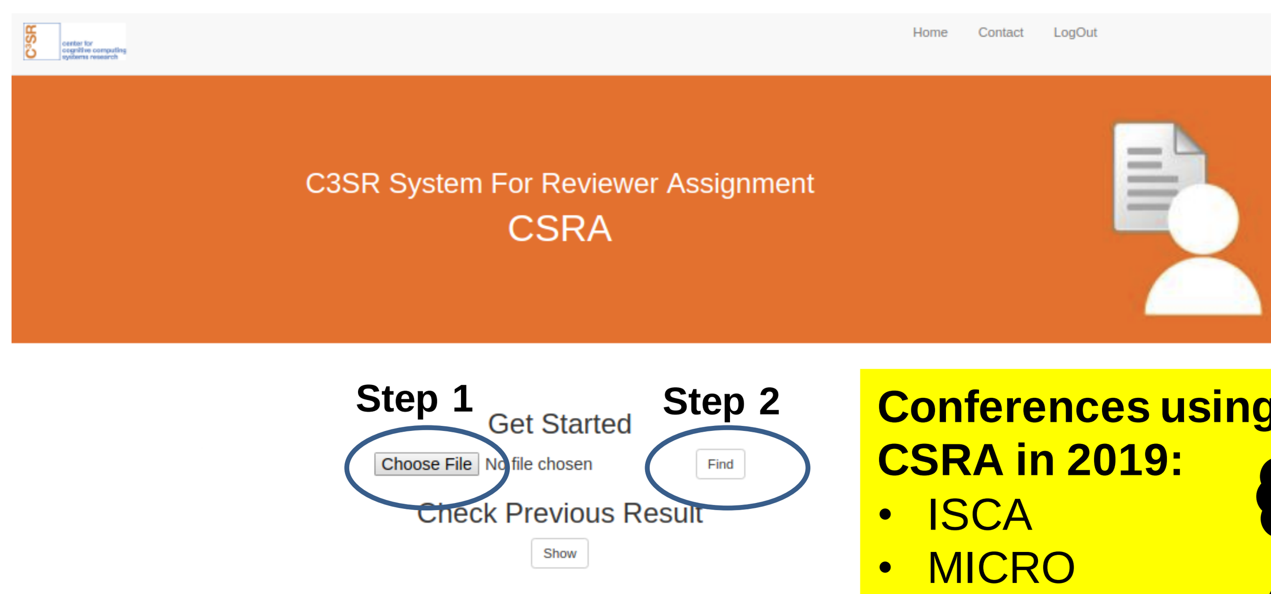
- Submission: ~150 papers published in NIPS 2006
- Reviewer pool: ~360 reviewers for NIPS 2006
- Annotation: Around 650 human judgments are available as groundtruth

Results

Method	P@5	P@10
Common Topic Model (Proposed)	**75.0**	**69.6**
Hidden Topic Model	59.3	51.7
LDA*	61.7	51.8
HDP	45.5	38.0
Doc2Vec	52.5	44.19
WMD	35.0	39.87

* Most of the state of the art systems including the ones mentioned in existing solutions use LDA

It is coming online!



Conferences using CSRA in 2019:

- ISCA
- MICRO
- HPCA

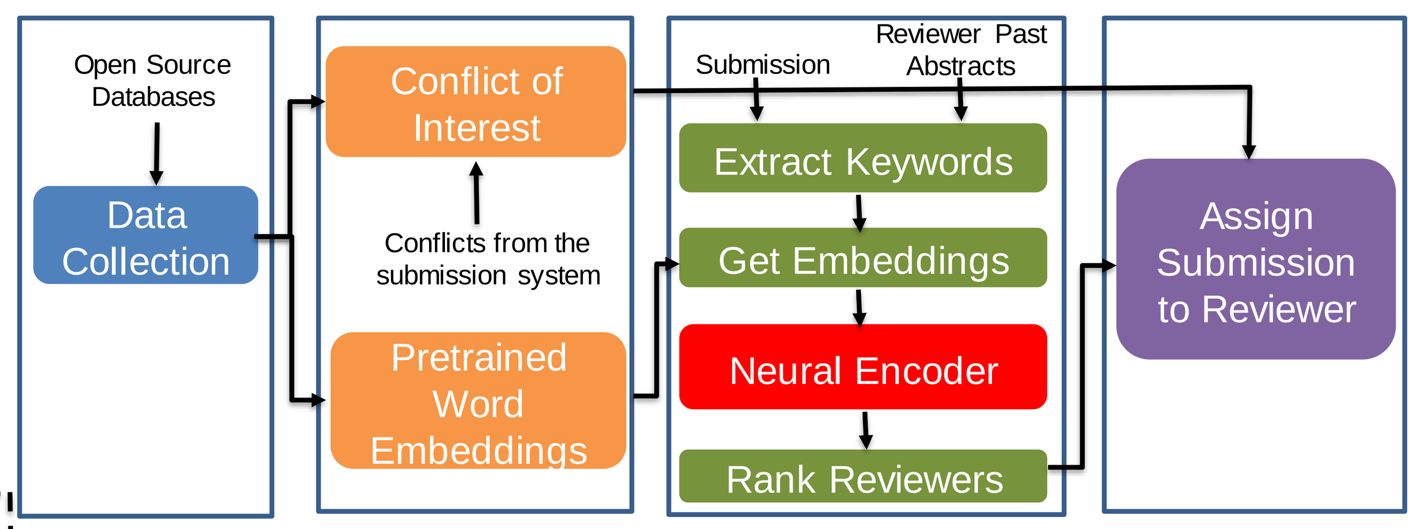
Conclusions

- This task heavily relies on extracting relevant keywords
- Use of embeddings instead of tokens is proven to be a robust way for this task to overcome vocabulary mismatch between submissions and reviewers
- The proposed common topic modeling shows a strong empirical performance over baselines

Future Work: Data collection

- Collection of new reviewer assignment datasets, which we have planned for the upcoming conferences including ISCA, MICRO and HPCA.
  - A large, randomly selected set of (reviewer, submission) pairs
  - A small but fully annotated set of (reviewer, submission) pairs

Future Work: System Development



- Evaluate the performance of more sophisticated neural encoders to profile reviewers
- Get supervision from readily available resources, by assuming that
  - Citation networks can provide unique signals of reviewers
  - One should be a good reviewer for his/her own papers.

References

[1] (Online) acl2019pcblog.fileli.unipi.it/?p=156
[2] David Mimno and Andrew McCallum. 2007. Expertise modeling for matching papers with reviewers. In *Proceedings of the 13th ACM SIGKDD international conference on Knowledge discovery and data mining* (KDD '07)
[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)
[4] L. Charlin, R. S. Zemel. The Toronto paper matchingsystem: an automated paper-reviewer assignmentsystem. (ICML) 2013

#AI Research Week

hosted by MIT-IBM Watson AI Lab

Detailed Contact Information: Omer Anjum, oanjum@illinois.edu