DeSCoVeR: Debiased Semantic Context Prior for Venue Recommendation

Sailaja Rajanala
Indian Institute of Technology Hyderabad
cs15resch11009@iith.ac.in

Arghya Pal
Harvard Medical School
arghyapal5@gmail.com

Manish Singh
Indian Institute of Technology Hyderabad
msingh@cse.iith.ac.in

Raphaël C.-W. Phan
Monash University, Malaysia
raphael.phan@monash.edu

KokSheik Wong
Monash University, Malaysia
wong.koksheik@monash.edu

ABSTRACT

We present a novel semantic context prior-based venue recommendation system that uses only the title and the abstract of a paper. Based on the intuition that the text in the title and abstract have both semantic and syntactic components, we demonstrate that joint training of a semantic feature extractor and syntactic feature extractor collaboratively leverages meaningful information that helps in recommending venues for paper publication. The proposed methodology that we call DeSCoVeR first elicits these semantic and syntactic features using a Neural Topic Model and text classifier, respectively. The model then executes a transfer learning optimization procedure to perform a contextual transfer between the feature distributions of the Neural Topic Model and the text classifier during the training phase. DeSCoVeR also mitigates the document-level label bias using a *Causal back-door path* criterion and a sentence-level keyword bias removal technique. Experiments on the DBLP dataset show that DeSCoVeR outperforms the state-of-the-art methods.

CCS CONCEPTS

• **Information systems** → **Probabilistic retrieval models**; • **Computing methodologies** → *Natural language generation*; *Learning settings*.

KEYWORDS

Document Classification, Topic Modeling, Joint Learning, Mutual Transfer, Causal Debiasing.

ACM Reference Format:
Sailaja Rajanala, Arghya Pal, Manish Singh, Raphaël C.-W. Phan, and KokSheik Wong. 2022. DeSCoVeR: Debiased Semantic Context Prior for Venue Recommendation. In *Proceedings of the 45th International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR '22), July 11–15, 2022, Madrid, Spain.* ACM, New York, NY, USA, 6 pages. https://doi.org/10.1145/3477495.3531877

SIGIR '22, July 11–15, 2022, Madrid, Spain

ACM ISBN 978-1-4503-8732-3/22/07…$15.00
https://doi.org/10.1145/3477495.3531877

1 INTRODUCTION

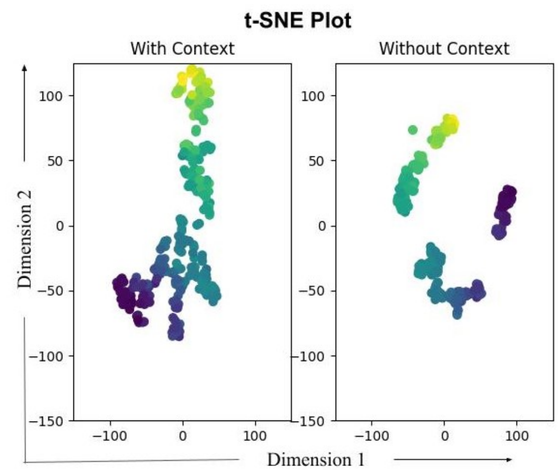

Figure 1: The visualization of the 16k papers (i.e., title and abstract of papers from a held-out test dataset) resulting from a t-SNE projection of the proposed DeSCoVeR indicates that the learned class neighborhood is locally continuous and globally compact.

A venue recommender system, $M(\cdot)$, entails the process of nominating a venue, v, or a set of appropriate venues for a given paper, i.e., $M(v|paper)$, such that the given paper has a high chance of acceptance when submitted to these venues [10, 41]. Researchers sought to submit their research findings to venues since they are a means to publish, register, and publicize one's work [25]. However, the acceptance process undertaken by the academic venues is highly competitive and thus asks for a befitting venue recommendation system that can align the properties of the paper with those of the venue. The task is challenging due to the increasing number of research topics and the venues, the dynamic nature of venue ratings, and the exponential growth of published articles in each domain.

Traditional venue recommender systems avail paper metadata such as the complete paper content, paper bibliography, author-publication history, and venue metadata such as venue location, rank, and topics. However, extensively collecting and processing data such as entire paper content and author publication history is costly and cumbersome. The technique of [1] recommends venues by extracting implicit feedback knowledge from the articles in users' Semantic Scholar[1] library. Works such as [44], [25], [8], [5], [23] and [22] estimate the venue, v, based on common interests of the all the authors, their co-authors and co-citers, i.e. $M(v|K)$ where K is the common interest knowledge base. On the other hand, a paper's full content or the minimalistic paper content i.e. its title, abstract and/or keywords is considered in [10, 17, 24, 33, 42], i.e. $M(v|paper)$. The works [11, 29, 35] and GraphConfRec proposed by [16] are a set of hybrid approaches that combine the benefits of minimalistic paper content with author publication history. Topical similarities between the papers using citation graphs are used in [6, 30].

We begin by asking what makes *a good venue recommender system?* To answer this question, we adopt the following two guiding principles:

Principle 1. $M(\cdot)$ *should leverage on semantic context priors in addition to its classification ability.*

Principle 2. *The learned* $M(\cdot)$ *must be free from any dataset bias such as label bias and keyword bias [31].*

Principle 1 helps to introduce the semantics of a text, i.e., features encapsulating the meaning of the text into the venue recommender together, thus combining the syntax and semantics of the language, while Principle 2 helps the model to mitigate any unintended language bias.

To this end, we propose DeSCoVER, aka Debiased Semantic Context Prior for Venue Recommendation, $M(\cdot)$, that operates *only* on the abstract and title of a paper and discovers the most suitable venue, v, i.e.$M(v|\mathbf{d}); \mathbf{d} : (abstract + title)$. Adopting the above two principles, we note in the left half of Fig. 1 that inducing the semantic prior regularizes the classifier to learn informative syntactic features and smoother class boundaries since venues are not homogeneous but are instead a mixture of various topics implying that the hard class boundaries are not very realistic.

First, DeSCoVeR performs a joint optimization that acquires the semantic context through a topic modeling network and leverages that in a syntactic document classifier. The contextual transfer boosts the performance of both the participating models by enriching the intermediate-level representations that share the backpropagation from both tasks. We note works [7, 14, 39] that recommend joint training of tasks on complementary contexts like emotion and sentiment classification. However, in contrast to these, we use joint training to leverage more complex modalities such as syntax and semantics. Thus, we propose a unified approach to make the most out of the available context in a short text, namely the title and the abstract of the paper. We hypothesize that this short text has a semantic component coming from its topical distribution and a syntactic component arising from phonology, morphology, syntax, and the pragmatics of the language.

Second, we perform a debiasing on a **trained** DeSCoVeR to mitigate the document-level label bias using a *back-door path* criterion [28], i.e. $M(v|do(\mathbf{d}))$, that appears especially due to an imbalance in venue label/category distribution. It also distills the word-level bias that appears due to excessive correlation between a set of words (e.g., the association of "chemical", "reaction", "solution" keywords with the chemical journals and conferences) using a keyword masking technique discussed in Sec. 3.

[1]https://www.semanticscholar.org/

Our contributions in this paper are as follows:

[Minimalistic Input] DeSCoVeR recommends a venue using only title and abstract;

[Collaborative Training] DeSCoVeR has a novel symbiotically-collaborative knowledge exchange mechanism for venue recommendation that outperforms all SOTAs, and;

[Debiasing] DeSCoVeR adopts a post-training mechanism to distill label and keyword bias.

2 RELATED WORK

We divide our discussion of related work into subsections that capture earlier efforts related to ours from different perspectives. We have segregated the related work based on venue classification approaches and the existing collaborative learning architectures followed by text debiasing techniques.

Venue Recommendation. An author's publication history and peer network disclose their research interests. For instance, [1] recommend venues by extracting implicit feedback knowledge from the articles in their Semantic Scholar library.

Works such as [44], [25], [8], [5], [23] and [22] are a group of collaborative recommender systems that estimate venue relevance based on common interests of all the authors, their co-authors and co-citers followed by optional post-filtering by venue metadata: namely, location, rank of the venue and the publishers. A paper's entire content is the most expressive part of the paper, thus proffering a class of content-based venue recommender systems such as [10] and [42]. On the contrary, the works [17, 24] and [33] merely rely on the minimalistic paper content i.e. its title, abstract and/or keywords. These works identify like content by clustering the topics of the short text using [4, 34]. The works [11, 29, 35] and GraphConfRec proposed by [16] are a set of hybrid approaches that combine the benefits of minimalistic paper content with author publication history. Citation graphs allude to topical similarities between the articles in the network [6]. [18] uses a class of direction-aware algorithms DARWR and DAKATZ on the citation graph for venue recommendation. CNAVER, by [30], first obtains seed venues from a citation-like graph to perform a RWR (Random Walk with Restart) on the venue-peer network to rank venues. [2] visualizes a heterogeneous graph containing authors, venues, and articles. To the best of our knowledge, previous works have only executed an ensemble approach to combine various textual features. However, our proposed method, DeSCoVer, is the only method that fortifies syntactic features of the text using its semantic features during training without the need for concatenation advocated by ensemble approaches.

Collaborative Learning. Collaborative learning methods adopt a mutual exchange or a mutual transfer approach to train the models in an amortized fashion. This is unlike the adversarial learning approach of the Generative Adversarial Network (GAN) [12] framework that leverages on a min-max game where the adversary tries to maximize the loss of the opponent while trying to minimize its own. The approach proposed in [15] engages two neural machine translators, a primal network (i.e., Source → Target) and the other its dual (i.e., Target → Source), in a learning game such that each of them acts as an adversary to the other. Transfer Cross Domain Recommendation (DDTCDR) [20] proposed a latent presentation

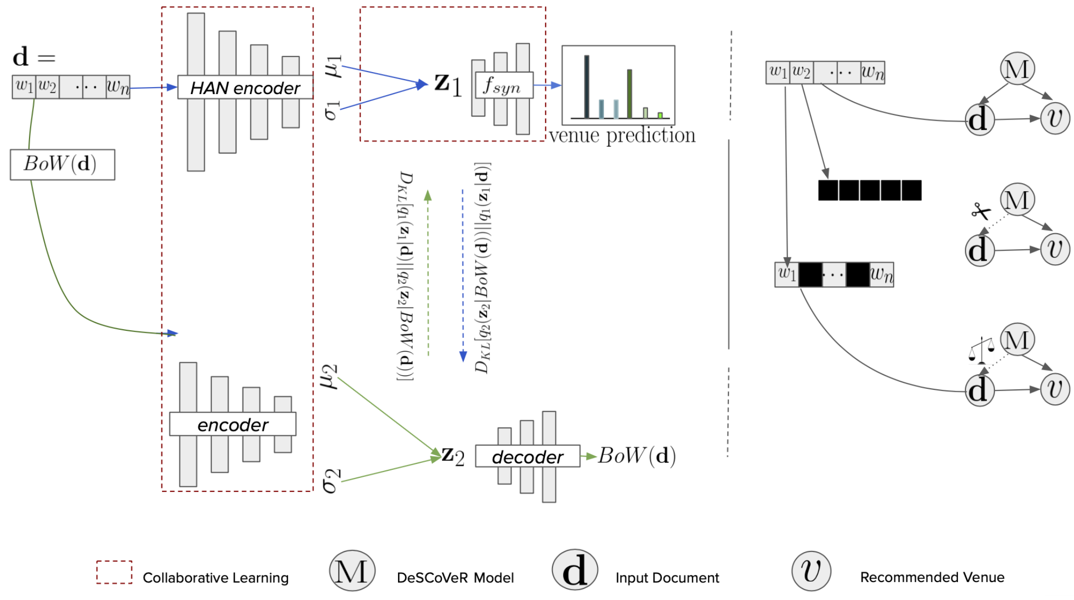

Figure 2: DeSCoVeR . (*Left*) The methodology discovers the most promising venue, v, given a paper/document, d, by leveraging on the semantic context prior within a classifier. The encoder of the semantic network provides a hidden representation, z_2, that collaboratively helps the hidden representation, z_1, of a syntactic HAN encoder. (*Right*) The DeSCoVeR performs debiasing to mitigate the document-level label bias using a *back-door path* criterion [28], i.e. $M(v|do(\mathbf{d}))$ removing the $M \rightarrow$ d link from the trained model M, also a sentence-level keyword bias.

for users that can be mutually shared across domains. Bidirectional Distillation (BD) of [19] extends the traditional teacher-student knowledge distillation method to a two-way knowledge transfer where not only a teacher teaches the student but also a student teaches the teacher. The cooperative Learning framework proposed by [3] enables cross-domain normalization of intermediate representations between the models. Two models working on the same task but different domains can exchange their knowledge at the attribute level. The Cooperative Training (CoT) framework of [21] modifies the popular adversarial, i.e., GAN framework [12, 13] to adopt a cooperative approach by replacing the adversarial network, i.e., discriminator, with a coordinator model called the mediator. The mediator estimates the combined density of the data and generated samples and uses the learned density to maximize the generative capabilities of the generator, thus replacing the min-max objective of the discriminator with a max-max objective function.

The methods mentioned above work with models operating on similar task objectives guiding each other. In this work, we show coordination between two different tasks, namely, a topic model and a text classifier.

Text Debiasing Dataset balancing, cleaning, and reweighing are the most adopted methods to solve the dataset imbalance problem [9, 32, 40]. However, they have been shown to be less effective [36] since train-dataset cleaning is a costly process, and it does not guarantee the non-susceptibility of the model when exposed to biased out-of-distribution data in the future. In this work, we adopt a counterfactual debiasing technique motivated by [31] to debias the model post-training. The debiasing technique is more generalized and is performed during the inference procedure and thus is less expensive.

3 DESCOVER METHODOLOGY

Definition. The DeSCoVeR , $M_\theta(v|\mathbf{d})$, is parametrized by θ, that can be optimized using cross entropy loss, i.e.

$$\mathcal{L}_{venue} = -\sum_{v=1}^{|\Omega|} y^v \log M_\theta(v|\mathbf{d}) \tag{1}$$

where y^v is a binary indicator variable on the true venue v, Ω is the set of all venues and the document $\mathbf{d} = \{w_1, w_2, \cdots, w_n\}$ has n words.

Backdoor adjustment using the causal graph. Prior to discussing the backdoor adjustment of the DeSCoVeR model, we discuss the causal graph [28] structure of the DeSCoVeR model. We view DeSCoVeR as a causal graph that describes a causal direction from $\{\mathbf{d}, M\}$ to v, i.e. $\{\mathbf{d}, M\} \rightarrow v$, shown in right hand side of Fig 2. We argue that the relation is contaminated by a spurious correlation $v \leftarrow M \rightarrow \mathbf{d}$. We take advantage of Pearl's *back-door adjustment* and decouple any such relation using the interventional $do(\cdot)$ operation, i.e. $M(v|do(\mathbf{d})) = M(v|\tilde{\mathbf{d}})$; where $\tilde{\mathbf{d}}$ is any counterfactual document that is not dependent on the dataset or $M(\cdot)$.

3.1 Training DeSCoVeR

In this section, we detail the training methodology of the DeSCoVeR model. Our methodology extracts the semantic and syntactic features and then optimizes the overall network using these features. Our methodology provides a special consideration for debiasing.

Syntactic feature extraction. We adopt a network that has two components, i.e. (i) the encoding network $q_1(\mathbf{d};\phi_1)$, and (ii) the classification network $p_1(v;\theta_1)$ that distils a document's representations at varying levels of granularity:

$$\mathcal{L}_{sy} = \min_{\phi_1,\theta_1} \mathbb{E}_{\mathbf{z}_1\sim q_1(\mathbf{d};\phi_1)}||v - p_1(\mathbf{z}_1;\theta_1)||_2^2 - D_{KL}[q_1(\mathbf{z}_1|\mathbf{d})||p(\mathbf{z}_1)] \tag{2}$$

We impose a Gaussian prior over the final document representation obtained by encoder $q_1(\mathbf{d};\phi_1)$ as: $\mathbf{z}_1 \in \mathbb{R}^K \sim \mathcal{N}(\mu_1, \sigma_1^2)$. $\mathbf{z}_1$ can be sampled as $\mathbf{z}_1 = \mu_1 + \sigma_1^2 \odot \epsilon$ where $\epsilon \in \mathcal{N}(0, I)$ and K is the number of topics in the document.
$\therefore$ $\mathbf{z}_1$ captures syntactic features.

Semantic feature extraction. We label a document's topics as its semantic features or conversely, its semantic context using the Neural Topic Model [26] with a variational semantic loss:

$$\mathcal{L}_{sm} = \min_{\phi_2,\theta_2} \mathbb{E}_{\mathbf{z}_2\sim q_2 1(BoW(\mathbf{d});\phi_2)}||BoW(\mathbf{d}) - p_2(\mathbf{z}_2;\theta_2)||_2^2 - D_{KL}[q_2(\mathbf{z}_2|BoW(\mathbf{d}))||p(\mathbf{z}_2)] \tag{3}$$

The encoder, $q_2 \sim (BoW(\mathbf{d});\phi_2)$, is a compression network that accepts a document $\mathbf{d}$'s bag-of-words $BoW(\cdot)$ and projects it onto a latent space $\mathbf{z} \in \mathbb{R}^K$ where K is the number of topics in the document. The decoder network, $p_\theta(\mathbf{z})$, reconstructs the bag-of-words representation $p(\mathbf{z})$.
$\therefore$ $\mathbf{z}_2$ captures semantic features.

Joint training for collaborative context transfer. In addition to Eqns. 2 and 3, the syntactic and semantic feature extractors are collaboratively trained via a joint training procedure, i.e.:

$$\begin{aligned}\mathcal{L}_{sy} &= \lambda_{sy} L_{sy} + (1-\lambda_{sy}) D_{KL}[q_1(\mathbf{z}_1|\mathbf{d})||q_2(\mathbf{z}_2|BoW(\mathbf{d}))] \\ \mathcal{L}_{sm} &= \lambda_{sm} L_{sm} + (1-\lambda_{sm}) D_{KL}[q_2(\mathbf{z}_2|BoW(\mathbf{d}))||q_1(\mathbf{z}_1|\mathbf{d})]\end{aligned} \tag{4}$$

where λ_{sy} and λ_{sm} are hyperparameters. Our final input to the network is $\mathbf{z}_2$ that has been enriched during the collaborative context transfer. This is unlike the traditional ensemble methods that use $z_{ens} = \mathbf{z}_1 \oplus \mathbf{z}_2$ and can mathematically be expressed as:

$$\mathcal{L}_{VR} = \min_{\substack{\phi_1,\theta_1\\ \phi_2,\theta_2}} \mathbb{E}_{\substack{\mathbf{z}_1\sim q_1(\mathbf{d};\phi_1)\\ \mathbf{z}_2\sim q_2(BoW(\mathbf{d});\phi_2)}} ||v - p_1(\mathbf{z}_{ens};\theta_1)||_2^2 \tag{5}$$

$\therefore$ Our final $M_\theta(v|\mathbf{d})$ is the $p_1(\mathbf{z}_1;\theta_1)$ that takes input from $q_1(\mathbf{d})$ that has been enriched by $q_2(BoW(\mathbf{d}))$ through contextual transfer from Eq. 4.

3.2 Debiasing of a trained DeSCoVeR

The debiasing technique eliminates the label and keyword biases present in the dataset or the trained model $M(\cdot)$. Label bias occurs when the number of training samples is disproportionate among the class labels [9] while the keyword bias can be seen when the decision of the classifier is correlated with the occurrence of a particular set of words.

Mitigation of label bias. We take the approach of [31] and provide a fully blindfolded document input $\tilde{\mathbf{d}} =< w_1, w_2, \cdots, w_n > =< MASK, MASK, \cdots, MASK >$ to both the encoders, i.e. $q_1(\tilde{\mathbf{d}})$ and $q_2(BoW(\tilde{\mathbf{d}}))$. Here, $MASK$ denotes a special token to mask a single word.
Label Bias Removal: The score $M(\tilde{v}|\tilde{\mathbf{d}})$ is subtracted from the $M(v|\mathbf{d})$, i.e. $M(v|\mathbf{d}) - M(\tilde{v}|\tilde{\mathbf{d}})$

Mitigation of keyword bias. Inspired by [28, 31] we deliberately expose a few words to elicit spurious correlations (e.g., "solution", "time", "reaction" to Chemistry related venues). Given a document $\mathbf{d}$ we perform a *text summarization* using the Jieba[2] tool to segregate the **content words** that provide classification clues from the **context words** that cause bias. The content words are *MASK*ed to get a biased document $\hat{\mathbf{d}}$, i.e., $\hat{\mathbf{d}} =< w_1, w_2, \cdots, w_n >=< MASK, w_2, \cdots, MASK >$. We can note that $\{MASK \forall w_i \in Content\}$. We pass $\hat{\mathbf{d}}$ on to both the encoders $q_1(\hat{\mathbf{d}})$ and $q_2(BoW(\hat{\mathbf{d}}))$.
Keyword Bias Removal: The score $M(\hat{v}|\hat{\mathbf{d}})$ is subtracted from the $M(v|\mathbf{d}) - M(\tilde{v}|\tilde{\mathbf{d}})$, i.e., $M(v|\mathbf{d}) - M(\tilde{v}|\tilde{\mathbf{d}}) - M(\hat{v}|\hat{\mathbf{d}})$.

4 DESCOVER EXPERIMENT

Dataset. We used a subset of the DBLP dataset of papers from 2009 to 2017 containing a paper id, author names, title, abstract, year of publication, and citation information. We removed papers that were missing titles or abstracts. We further filtered the dataset such that every venue had a minimum of 100 papers. After processing, we were left with 1,68,103 papers, 223 venues, 2,22,651 unique authors, an average of 155 words per short text, and 1062 characters.
Baseline Methods. Due to the novelty and expressibility of our work, we have two types of baselines:

- Architectural Baselines, BERT [38], HAN [43], 4-layered Bi-LSTM, TextCNN with kernel sizes from the set $\in \{2, 3, 4, 5\}$, 32 filters, an NTM based semantic classifier, a feature level ensemble of NTM and Classifier.
- Literary Baselines consisting of Author-Collaborative Filtering, LDA-Content Based, Author Neighborhood [22] and finally a Cavnar Trenkle text classification [24] baseline.

We evaluate the baselines based on accuracy, precision, and F1 score, and the results have been tabulated in Table 1. We chose accuracy as the evaluation metric since it was commonly used by SOTA methods such as HAN, BERT, based text classifiers.

Our Method. Our syntactic network, $p_1(q_1(\mathbf{d}))$, is comprised of the Hierarchical Attention Network (HAN) [43] as the encoder extended by a three-level decoder followed by a fully connected classifying layer. The semantic network, $p_2(q_2(BoW(\mathbf{d})))$, is the Neural Topic Models (NTM) [26, 27, 37].

Training Details. Warm up stage: We train NTM for 10 epochs and the document classifier for 2 epochs during which the models gather their respective contexts. Joint training: following warmup, we train the two models in a mutual exchange mode for 30 epochs where the semantic context from NTM is passed on to the classifier and vice-versa. We have a batch size of 32; the classifier learning

[2] https://github.com/fxsjy/jieba

Method	Acc(↑)	Precision(↑)	F1(↑)
Collaborative Filtering	0.18	0.07	0.08
Content Based	0.07	0.02	0.03
Author Neighbourhood [22]	0.13	0.04	0.04
Cavnar Trenkle [24]	0.09	0.03	0.03
LSTM classifier	0.10	0.08	0.08
CNN classifier	0.09	0.06	0.04
BERT classifier	0.30	0.25	0.24
Syntactic Classifier (HAN)	0.33	0.31	0.30
Semantic Classifier (NTM)	0.28	0.25	0.25
Ensemble	0.34	0.32	0.31
Biased DeSCoVeR Title+Abstract	**0.37**	**0.34**	**0.34**

Table 1: Biased DeSCoVeR vs Biased Baselines Acc, Precision and F1 (averaged over five runs), higher (↑) is better.

Method	Acc	Precision	F1
DeSCoVeR Title+Abstract - (Label+Keyword) Bias	**0.382**	**0.353**	**0.351**
DeSCoVeR Title+Abstract - Label Bias	0.374	0.341	0.342
DeSCoVeR Title+Abstract - Keyword Bias	0.378	0.344	0.343
DeSCoVeR Abstract - Label Bias	0.295	0.261	0.258
DeSCoVeR Abstract - Keyword Bias	0.301	0.261	0.258
DeSCoVeR Abstract - (Label+Keyword) Bias	0.31	0.265	0.262
HAN - Keyword Bias	0.336	0.310	0.308
HAN - Label Bias	0.332	0.310	0.308
HAN - (Label + Keyword) Bias	0.342	0.320	0.321

Table 2: Comparison of debiasing on methods. Debiasing improves Acc, Precision, and F1 (averaged over five runs each) of DeSCoVeR and HAN. We note similar improvements in other methods (not reported here for conciseness of space).

rate is 0.001, NTM learning rate is 0.005. An evaluation of NTM by varying topics $K = [5, 80]$ gave the highest topic coherence at $K = 50$. The $\lambda_{sem} = 0.2$, $\lambda_{syn} = 0.6$, are chosen after a beam-search. We used python-3.7.3, PyTorch-1.0.1, and NVIDIA P1000 GPU.

5 RESULTS AND DISCUSSION

Biased DeSCoVeR vs. Biased Baselines. As discussed in Sec. 3.1, we first train DeSCoVeR *without* bias correction and compared the model with the baseline methods such as HAN, NTM, author neighbourhood [22], collaborative filtering [1], Content Based [11] and Cavnar Trenkle [24] as shown in Table 1. We see that HAN, BERT, and the Ensemble methods are the top contenders alongside our proposed method, DeSCoVeR. We may attribute this performance boost to the smoother data space learned by DeSCoVeR using the semantic and syntactic features, a proof of which is shown in Fig. 1 of Sec. 1.

Method	Acc(↑)	Precision(↑)	F1(↑)
DeSCoVeR Title	0.28	0.25	0.25
DeSCoVeR Abstract	0.29	0.26	0.25
DeSCoVeR Title+Abstract	**0.37**	**0.34**	**0.34**
DeSCoVeR Area	0.61	0.59	0.57
DeSCoVeR Subarea	0.49	0.46	0.45

Table 3: Ablation Studies of Biased DeSCoVeR

Ablation Studies. The ablation study was conducted w.r.t. the role of title and abstract on venue recommendation. We also study the performance of DeSCoVeR in the area and subarea levels, i.e., 20 areas and 77 subareas, by grouping similar venues together using labels from WikiCFP[3], see Table 3. Area recommendation is accomplished by predicting the right field of study for the paper and subarea prediction is the problem of identifying the more specific field of study in the broader area.

Effect of Debiasing. We show in Table 2 the efficacy of mitigating the label and keyword debiasing that we discussed in Sec. 3.2. Owing to space constraints, we only report in Table 2 the debiasing experiments conducted on HAN and the proposed DeSCoVeR. However, we note an improvement over every baseline methods after bias correction. In general we observed that the keyword bias to be more profound. This can result from the amplification of keyword bias when learning hierarchical attentions at word, sentence and paragraph level during classification. To elaborate, we note $\sim 1\%$ improvement on accuracy for DeSCoVeR that considers title and abstract as input. Similar trends are observed for other combinations of the DeSCoVeR model.

6 CONCLUSION

We showed a novel semantic context prior-based venue recommendation system that uses only the title and the abstract of a paper. Although we tried to debias the model using known debiasing techniques, the debiasing is domain-specific and the results may vary across domains. DeSCoVeR offers promising solutions for cold start problems since it can recommend venues using just the title and the abstract without asking for a user's past history, which suits the double-blind review process.

[3]http://www.wikicfp.com/cfp/

REFERENCES

[1] Hamed Alhoori and Richard Furuta. 2017. Recommendation of scholarly venues based on dynamic user interests. *Journal of Informetrics* 11, 2 (2017), 553–563.

[2] Abdulrhman M Alshareef, Mohammed F Alhamid, and Abdulmotaleb El Saddik. 2019. Academic venue recommendations based on similarity learning of an extended nearby citation network. *IEEE Access* 7 (2019), 38813–38825.

[3] Tanmay Batra and Devi Parikh. 2017. Cooperative learning with visual attributes. *arXiv preprint arXiv:1705.05512* (2017).

[4] David M Blei, Andrew Y Ng, and Michael I Jordan. 2003. Latent dirichlet allocation. *Journal of machine Learning research* 3, Jan (2003), 993–1022.

[5] Imen Boukhris and Raouia Ayachi. 2014. A novel personalized academic venue hybrid recommender. In *2014 IEEE 15th International Symposium on Computational Intelligence and Informatics (CINTI)*. IEEE, CINTI, 465–470.

[6] Cornelia Caragea and Corina Florescu. 2018. Venue Classification of Research Papers in Scholarly Digital Libraries. In *International Conference on Theory and Practice of Digital Libraries*. Springer, 129–136.

[7] Charlotte Caucheteux, Alexandre Gramfort, and Jean-Remi King. 2021. Disentangling syntax and semantics in the brain with deep networks. In *International Conference on Machine Learning*. PMLR, 1336–1348.

[8] Zhen Chen, Feng Xia, Huizhen Jiang, Haifeng Liu, and Jun Zhang. 2015. AVER: random walk based academic venue recommendation. In *Proceedings of the 24th International Conference on World Wide Web*. ACM, 579–584.

[9] Lucas Dixon, John Li, Jeffrey Sorensen, Nithum Thain, and Lucy Vasserman. 2018. Measuring and mitigating unintended bias in text classification. In *Proceedings of the 2018 AAAI/ACM Conference on AI, Ethics, and Society*. 67–73.

[10] Tirthankar Ghosal, Ashish Raj, Asif Ekbal, Sriparna Saha, and Pushpak Bhattacharyya. 2019. A Deep Multimodal Investigation To Determine the Appropriateness of Scholarly Submissions. In *2019 ACM/IEEE Joint Conference on Digital Libraries (JCDL)*. IEEE, 227–236.

[11] Tirthankar Ghosal, Ravi Sonam, Asif Ekbal, Sriparna Saha, and Pushpak Bhattacharyya. 2019. Is the Paper Within Scope? Are You Fishing in the Right Pond?. In *2019 ACM/IEEE Joint Conference on Digital Libraries (JCDL)*. IEEE, 237–240.

[12] Ian Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio. 2014. Generative adversarial nets. *Advances in neural information processing systems* 27 (2014).

[13] Ian Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio. 2020. Generative adversarial networks. *Commun. ACM* 63, 11 (2020), 139–144.

[14] Serhii Havrylov, Germán Kruszewski, and Armand Joulin. 2019. Cooperative learning of disjoint syntax and semantics. *arXiv preprint arXiv:1902.09393* (2019).

[15] Di He, Yingce Xia, Tao Qin, Liwei Wang, Nenghai Yu, Tie-Yan Liu, and Wei-Ying Ma. 2016. Dual learning for machine translation. *Advances in neural information processing systems* 29 (2016), 820–828.

[16] Andreea Iana and Heiko Paulheim. 2021. GraphConfRec: A Graph Neural Network-Based Conference Recommender System. *arXiv preprint arXiv:2106.12340* (2021).

[17] Konstantin Kobs, Tobias Koopmann, Albin Zehe, David Fernes, Philipp Krop, and Andreas Hotho. 2020. Where to Submit? Helping Researchers to Choose the Right Venue. In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing: Findings*. 878–883.

[18] Onur Küçüktunç, Erik Saule, Kamer Kaya, and Ümit V Çatalyürek. 2012. Recommendation on academic networks using direction aware citation analysis. *arXiv preprint arXiv:1205.1143* (2012).

[19] Wonbin Kweon, SeongKu Kang, and Hwanjo Yu. 2021. Bidirectional Distillation for Top-K Recommender System. In *Proceedings of the Web Conference 2021*. 3861–3871.

[20] Pan Li and Alexander Tuzhilin. 2020. Ddtcdr: Deep dual transfer cross domain recommendation. In *Proceedings of the 13th International Conference on Web Search and Data Mining*. 331–339.

[21] Sidi Lu, Lantao Yu, Siyuan Feng, Yaoming Zhu, and Weinan Zhang. 2019. Cot: Cooperative training for generative modeling of discrete data. In *International Conference on Machine Learning*. PMLR, 4164–4172.

[22] Hiep Luong, Tin Huynh, Susan Gauch, Loc Do, and Kiem Hoang. 2012. Publication venue recommendation using author network's publication history. *Intelligent Information and Database Systems* (2012), 426–435.

[23] Hiep Phuc Luong, Tin Huynh, Susan Gauch, and Kiem Hoang. 2012. Exploiting Social Networks for Publication Venue Recommendations.. In *Kdir*. 239–245.

[24] Eric Medvet, Alberto Bartoli, and Giulio Piccinin. 2014. Publication venue recommendation based on paper abstract. In *Tools with Artificial Intelligence (ICTAI), 2014 IEEE 26th International Conference on*. IEEE, 1004–1010.

[25] Nour Mhirsi and Imen Boukhris. 2017. Exploring location and ranking for academic venue recommendation. In *International Conference on Intelligent Systems Design and Applications*. Springer, 83–91.

[26] Yishu Miao, Edward Grefenstette, and Phil Blunsom. 2017. Discovering discrete latent topics with neural variational inference. In *International Conference on Machine Learning*. PMLR, 2410–2419.

[27] Yishu Miao, Lei Yu, and Phil Blunsom. 2016. Neural variational inference for text processing. In *International conference on machine learning*. PMLR, 1727–1736.

[28] Judea Pearl. 2009. *Causality*. Cambridge university press.

[29] Tribikram Pradhan, Abhinav Gupta, and Sukomal Pal. 2020. HASVRec: A modularized Hierarchical Attention-based Scholarly Venue Recommender system. *Knowledge-Based Systems* 204 (2020), 106181.

[30] Tribikram Pradhan and Sukomal Pal. 2020. CNAVER: A content and network-based academic VEnue recommender system. *Knowledge-Based Systems* 189 (2020), 105092.

[31] Chen Qian, Fuli Feng, Lijie Wen, Chunping Ma, and Pengjun Xie. 2021. Counterfactual Inference for Text Classification Debiasing. In *Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (Volume 1: Long Papers)*. 5434–5445.

[32] Colin Raffel, Noam Shazeer, Adam Roberts, Katherine Lee, Sharan Narang, Michael Matena, Yanqi Zhou, Wei Li, and Peter J Liu. 2019. Exploring the limits of transfer learning with a unified text-to-text transformer. *arXiv preprint arXiv:1910.10683* (2019).

[33] Sailaja Rajanala and Manish Singh. 2020. FLY: Venue Recommendation using Limited Context. In *2020 IEEE 32nd International Conference on Tools with Artificial Intelligence (ICTAI)*. IEEE, 200–204.

[34] Daniel Ramage, David Hall, Ramesh Nallapati, and Christopher D Manning. 2009. Labeled LDA: A supervised topic model for credit attribution in multi-labeled corpora. In *Proceedings of the 2009 Conference on Empirical Methods in Natural Language Processing: Volume 1-Volume 1*. Association for Computational Linguistics, ACL, 248–256.

[35] Ramin Safa, SeyedAbolghasem Mirroshandel, Soroush Javadi, and Mohammad Azizi. 2018. Venue Recommendation Based on Paper's Title and Co-authors Network. *Journal of Information Systems and Telecommunication* 6, 1 (2018), 33–40.

[36] Timo Schick, Sahana Udupa, and Hinrich Schütze. 2021. Self-diagnosis and self-debiasing: A proposal for reducing corpus-based bias in nlp. *Transactions of the Association for Computational Linguistics* 9 (2021), 1408–1424.

[37] Akash Srivastava and Charles Sutton. 2017. Autoencoding variational inference for topic models. *arXiv preprint arXiv:1703.01488* (2017).

[38] Thomas Wolf, Lysandre Debut, Victor Sanh, Julien Chaumond, Clement Delangue, Anthony Moi, Pierric Cistac, Tim Rault, Rémi Louf, Morgan Funtowicz, Joe Davison, Sam Shleifer, Patrick von Platen, Clara Ma, Yacine Jernite, Julien Plu, Canwen Xu, Teven Le Scao, Sylvain Gugger, Mariama Drame, Quentin Lhoest, and Alexander M. Rush. 2020. Transformers: State-of-the-Art Natural Language Processing. In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing: System Demonstrations*. Association for Computational Linguistics, Online, 38–45. https://www.aclweb.org/anthology/2020.emnlp-demos.6

[39] Lili Yang, Chunping Li, Qiang Ding, and Li Li. 2013. Combining lexical and semantic features for short text classification. *Procedia Computer Science* 22 (2013), 78–86.

[40] Xu Yang, Hanwang Zhang, and Jianfei Cai. 2021. Deconfounded Image Captioning: A Causal Retrospect. *IEEE Transactions on Pattern Analysis & Machine Intelligence* 01 (2021), 1–1.

[41] Zaihan Yang and Brian D. Davison. 2012. Venue Recommendation: Submitting Your Paper with Style. In *2012 11th International Conference on Machine Learning and Applications*, Vol. 1. ICML, 681–686. https://doi.org/10.1109/ICMLA.2012.127

[42] Zaihan Yang and Brian D Davison. 2012. Venue recommendation: Submitting your paper with style. In *Machine learning and applications (ICMLA), 2012 11th international conference on*, Vol. 1. IEEE, 681–686.

[43] Zichao Yang, Diyi Yang, Chris Dyer, Xiaodong He, Alex Smola, and Eduard Hovy. 2016. Hierarchical attention networks for document classification. In *Proceedings of the 2016 conference of the North American chapter of the association for computational linguistics: human language technologies*. NAACL, 1480–1489.

[44] Abir Zawali and Imen Boukhris. 2018. A group recommender system for academic venue personalization. In *International Conference on Intelligent Systems Design and Applications*. Springer, 597–606.