

# TA-DA: Topic-Aware Domain Adaptation for Scientific Keyphrase Identification and Classification (Student Abstract)

Răzvan-Alexandru Smădu<sup>1</sup>, George-Eduard Zaharia<sup>1</sup>, Andrei-Marius Avram<sup>1</sup>,  
Dumitru-Clementin Cercel<sup>1</sup>, Mihai Dascalu<sup>1</sup>, Florin Pop<sup>1,2</sup>

<sup>1</sup>Faculty of Automatic Control and Computers, University Politehnica of Bucharest

<sup>2</sup>National Institute for Research and Development in Informatics - ICI Bucharest, Romania

{razvan.smadu, george.zaharia0806, andrei\_marius.avram}@stud.acs.upb.ro

{dumitru.cercel, mihai.dascalu, florin.pop}@upb.ro

## Abstract

Keyphrase identification and classification is a Natural Language Processing and Information Retrieval task that involves extracting relevant groups of words from a given text related to the main topic. In this work, we focus on extracting keyphrases from scientific documents. We introduce TA-DA, a Topic-Aware Domain Adaptation framework for keyphrase extraction that integrates Multi-Task Learning with Adversarial Training and Domain Adaptation. Our approach improves performance over baseline models by up to 5% in the exact match of the F1-score.

## Introduction

Scientific Keyphrase Identification and Classification (SKIC) refers to the labeling of relevant words from a given input scientific document, which the SemEval-2017 workshop (Augenstein et al. 2017) introduced in Task 10. This task is related to Named Entity Recognition (NER), where the aim is to extract and classify the named entities from a given set of documents. SKIC can be more challenging than NER problems due to the lack of available annotated scientific publications (Augenstein et al. 2017).

This work focuses on using domain adaptation through adversarial training and adversarial examples to address SKIC. We introduce a Topic-Aware Domain Adaptive (TA-DA) deep neural network framework that incorporates multi-task learning and adversarial learning. Figure 1 briefly presents the training procedure. Experiments showed that each component improves the performance of our model on all considered datasets. We summarize our main contributions as follows:

- We propose a novel neural network architecture integrating multi-task learning, adversarial training, domain adaptation, and topic modeling for jointly addressing SKIC tasks;
- Different techniques for latent representations required for topic modeling (i.e., LSA, NMF, K-Means, LDA) are considered, and LDA provided the most promising overall results;

Copyright © 2023, Association for the Advancement of Artificial Intelligence (www.aaai.org). All rights reserved.

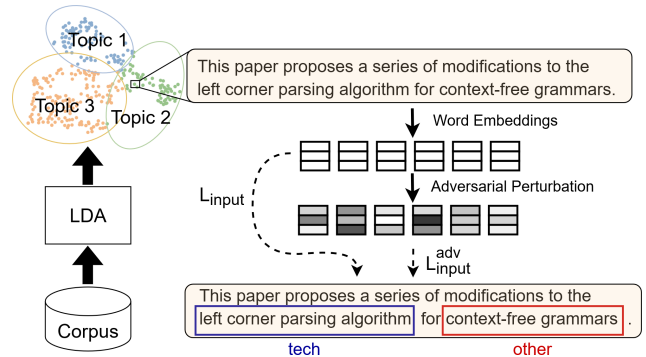


Figure 1: The training procedure for TA-DA employing domain adaptation and adversarial training.

- We show that domain adaptation improves performance compared to baselines, while adding adversarial examples further increases the F1-scores.

## Methodology

**Data Representation.** We cast the problem of SKIC to a tagging task. Given a document  $d = \{w_1, \dots, w_n\}$  with  $n$  words, our goal is to identify and classify keyphrases by outputting a sequence  $\{y_1, \dots, y_n\}$ , where each  $y_i$  is a class from the BIO schema (Lample et al. 2016).

**LDA for Topic Modeling.** Our intuition is to extract latent topics related to the domains and combine them with the labels of examples. Latent Dirichlet Allocation (LDA) (Blei, Ng, and Jordan 2003) is a frequently employed unsupervised method to determine latent topics from a corpus. It associates each document with a mixture of topics as probability distributions of words that are thematically related based on tight co-occurrence.

**Neural Network Architecture.** We propose a neural network architecture starting with a pre-trained SciBERT (Beltagy, Lo, and Cohan 2019) that generates contextualized embeddings. To further enhance the latent space, this representation is fed into a stack of BiLSTM layers, whose outputs are concatenated along with the SciBERT's output. We use multi-task learning to incorporate branches for each task that share the feature representation. After the feature extractor,

fully connected layers are used for keyphrase identification (KI) and classification (KC) tasks. For KI, the output is a linear-chain CRF, which models the probability of having the output label given the latent representation. For KC, a softmax activation function is considered to model a probability distribution across classes. We incorporate adversarial training by including a domain discriminator linked via a gradient reversal layer (Ganin and Lempitsky 2015) to the feature extractor. The output is computed using a softmax layer that models a probability distribution across domains, i.e. the topics extracted using LDA.

**Adversarial Learning.** We perform domain adaptation through adversarial training (Ganin and Lempitsky 2015). In this configuration, adversarial training seeks the saddle point that minimizes the expected likelihood of the loss associated with the label predictor while maximizing the likelihood for the discriminator. Additionally, we employ adversarial examples via Fast Gradient Sign Method (Goodfellow, Shlens, and Szegedy 2015) that adds a small value to the input (in our case, the latent space generated by SciBERT) proportional to the gradient in order to create more robust models. The new input is defined as follows:

$$x_{adv} = x + \epsilon \text{sign}(\nabla_x L(\theta, x, y)) \quad (1)$$

where  $\epsilon$  is the perturbation factor.

**Optimization.** The optimization process involves minimizing the negative log-likelihood of the CRF output and the categorical cross-entropy of the softmax outputs. The total loss is described by:

$$L_{total} = L_{tag} + L_{class} + \lambda L_{da} \quad (2)$$

where  $L_{tag}$  is the KI loss,  $L_{class}$  is the KC loss,  $\lambda$  controls the degree of enforced domain adaptation, and  $L_{da}$  is the domain adaptation loss. In the case of using adversarial examples, the loss  $L_{total}^{adv}$  is similar to (2), and both losses for each pass are added during the training step.

## Results

**Datasets.** We evaluate our approach on three datasets: SemEval-2017 Task 10 Dataset (Augenstein et al. 2017), SciERC (SciIE) Dataset (Luan et al. 2018), and ACL RD-TEC 2.0 Dataset (QasemiZadeh and Schumann 2016).

**Experimental Results.** The results are shown in Table 1. Our approach (i.e., DA-Adv) obtains the best or second-best scores concerning the baselines. Although domain adaptation may hinder performance by 1 to 2%, we observe that adversarial examples enhance the model’s performance in every scenario. Additionally, including LSA instead of LDA for topic modeling (i.e., DA-LSA), we observe similar performances, albeit on average, LDA performs better. Conversely, NMF (i.e., DA-NMF) and K-Means (i.e., DA-KM) hinder performances when employed as topic models. Compared with LSTM-CRF (Lample et al. 2016), our approach is comparable or performs better. Also, our model has considerably fewer trainable parameters (2.4M) than SciBERT (Beltagy, Lo, and Cohan 2019) (109M) and improved results when compared with similar approaches (i.e., MTL-LSTM (Augenstein and Søgaard 2017)).

Model	SemEval		SciIE		ACL RD-TEC	
	KI	KIC	KI	KIC	KI	KIC
DA-Adv	53.6	<u>42.6</u>	78.4	65.9	<u>82.5</u>	<b>70.0</b>
DA-LSA	53.2	41.5	<u>78.9</u>	<u>66.8</u>	<b>82.7</b>	<u>69.5</u>
DA-NMF	54.0	41.1	76.6	63.6	81.6	69.3
DA-KM	49.7	38.0	77.7	65.3	81.5	69.0
LSTM-CRF	58.6	33.7	-	-	74.4	35.5
MTL-LSTM	<b>67.7</b>	38.0	72.3	58.1	81.9	58.5
SciBERT	<u>66.7</u>	<b>48.2</b>	<b>81.0</b>	<b>67.4</b>	79.8	65.1

Table 1: F1-scores on the test sets (the highest scores are bolded, and the second-highest scores are underlined).

## Conclusions

We proposed TA-DA, a topic-aware domain adaptation approach for tackling SKIC. We conducted multiple experiments and analyzed the performance impact of adversarial learning and topic modeling. We observed that the adversarially trained models outperform other models on one dataset, while considerable improvements are achieved in contrast to LSTM-CRF-based models. Moreover, we showed that domain adaptation and adversarial examples improved results over baselines. In future work, we aim to study a hierarchical representation of the input to increase robustness further.

## Acknowledgements

The research has been funded by the University Politehnica of Bucharest through the PubArt program.

## References

- Augenstein, I.; Das, M.; Riedel, S.; Vikraman, L.; and McCallum, A. 2017. SemEval 2017 Task 10: ScienceIE - Extracting Keyphrases and Relations from Scientific Publications. In *SemEval*, 546–555.
- Augenstein, I.; and Søgaard, A. 2017. Multi-Task Learning of Keyphrase Boundary Classification. In *ACL*, 341–346.
- Beltagy, I.; Lo, K.; and Cohan, A. 2019. SciBERT: A Pretrained Language Model for Scientific Text. In *EMNLP-IJCNLP*, 3615–3620.
- Blei, D. M.; Ng, A. Y.; and Jordan, M. I. 2003. Latent dirichlet allocation. *JMLR*, 3: 993–1022.
- Ganin, Y.; and Lempitsky, V. 2015. Unsupervised domain adaptation by backpropagation. In *ICML*, 1180–1189.
- Goodfellow, I. J.; Shlens, J.; and Szegedy, C. 2015. Explaining and Harnessing Adversarial Examples. In *ICLR*.
- Lample, G.; Ballesteros, M.; Subramanian, S.; Kawakami, K.; and Dyer, C. 2016. Neural Architectures for Named Entity Recognition. In *NAACL*, 260–270.
- Luan, Y.; He, L.; Ostendorf, M.; and Hajishirzi, H. 2018. Multi-Task Identification of Entities, Relations, and Coreference for Scientific Knowledge Graph Construction. In *EMNLP*, 3219–3232.
- QasemiZadeh, B.; and Schumann, A.-K. 2016. The ACL RD-TEC 2.0: A language resource for evaluating term extraction and entity recognition methods. In *LREC*, 1862–1868.