Model	Inspec		Krapivin		NUS		SemEval		KP20k	
	F_1@5	F_1@10	F_1@5	F_1@10	F_1@5	F_1@10	F_1@5	F_1@10	F_1@5	F_1@10
TF-IDF	0.221	0.313	0.129	0.160	0.136	0.184	0.128	0.194	0.102	0.162
TextRank	0.223	0.281	0.189	0.162	0.195	0.196	0.176	0.187	0.175	0.147
Maui	0.040	0.042	0.249	0.216	0.249	0.268	0.044	0.039	0.270	0.230
CopyRNN	0.278	0.342	0.311	0.266	0.334	0.326	0.293	0.304	0.333	0.262
CopyCNN	0.285	0.346	0.314	0.272	0.342	0.330	0.295	0.308	0.351	0.288
TG-Net	**0.315**	**0.381**	**0.349**	**0.295**	**0.406**	**0.370**	**0.318**	**0.322**	**0.372**	**0.315**
% gain	10.5%	10.1%	11.1%	8.5%	18.7%	12.1%	7.8%	4.5%	6.0%	9.4%

Table 3: Present keyphrase predicting results on all test datasets. "% gain" is the improvement gain over CopyCNN.

Model	Inspec		Krapivin		NUS		SemEval		KP20k	
	R@10	R@50	R@10	R@50	R@10	R@50	R@10	R@50	R@10	R@50
CopyRNN	0.047	0.100	0.113	0.202	0.058	0.116	0.043	0.067	0.125	0.211
CopyCNN	0.050	0.107	0.119	0.205	0.062	0.120	0.044	0.074	0.147	0.225
TG-Net	**0.063**	**0.115**	**0.146**	**0.253**	**0.075**	**0.137**	**0.045**	**0.076**	**0.156**	**0.268**
% gain	26.0%	7.5%	22.7%	23.4%	21.0%	14.2%	2.3%	2.7%	6.1%	19.1%

Table 4: Absent keyphrase predicting results on all test datasets. "% gain" is the improvement gain over CopyCNN.

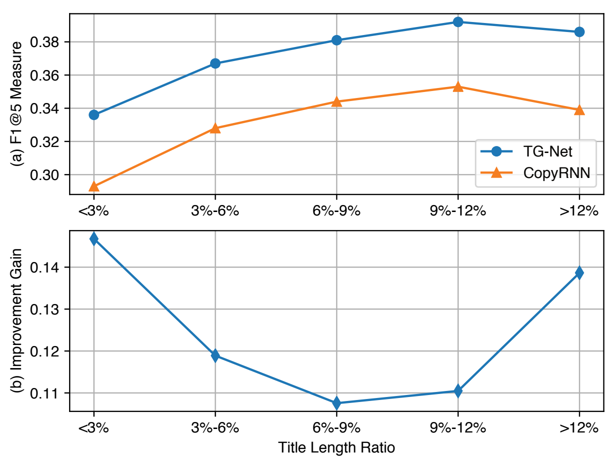

Figure 3: Present keyphrase predicting ability (F1@5 measure) on various title length ratios.

Keyphrase Predicting on Various Title Length Ratios

To find out how our title incorporation influences the prediction ability, we compare the keyphrase predicting ability of two RNN-based models (i.e., our model and CopyRNN) on different title length ratios. The title length ratio is defined as the title length over the context length. This analysis is based on the **KP20k** testing dataset. In view of the title length ratio, we preprocess the testing set into five groups (i.e., <3%, 3%-6%, 6%-9%, 9%-12% and >12%). Then, the present keyphrase prediction results (F1@5 measure) and the improvement gain on each group are depicted in Figure 3.

In Figure 3(a), we notice that both CopyRNN and our TG-Net model generally perform better when the title length ratio is higher. One possible explanation is that when the title is long, it conveys substantial salient information of the abstract. Therefore, the chance for the models to attend to the core information is enhanced, which leads to the observed situation. This figure also shows that both TG-Net and CopyRNN get worse performance on >12% group than 9%-12% group. The main reason is that there exist some data with a short abstract in >12% group, which leads to the lack of enough context information for correctly generating all keyphrases.

In Figure 3(b), we find that our TG-Net consistently improves the performance with a large margin on five testing groups, which again indicates the effectiveness of our model. In a finer perspective, we note that the improvement gain is higher on the lowest (i.e., <3%) and the highest (i.e., >12%) title length ratio groups. In >12% group, the title plays a more important role than in other groups, and consequently our model benefits more by not only explicitly emphasizing the title information itself, but also utilizing it to guide the encoding of information in the main body. As for <3% group, the effect of such a short title is small on the latter part of the context in CopyRNN because of the long distance. However, our model explicitly employs the title to guide the encoding of each context word regardless of the distance, which utilizes the title information much more sufficiently. Consequently, our model achieves much higher improvement in this group. While we only display the results of present keyphrase prediction, the absent keyphrase predicting task gets the similar results.

Ablation Study

We also perform an ablation study on **Krapivin** for better understanding the contributions of the main parts of our model. For a comprehensive comparison, we conduct this study on both present keyphrase prediction and absent keyphrase prediction.

Title: Exponential stability of switched stochastic delay systems with **non-linear uncertainties**
Abstract: This article considers the robust **exponential stability** of uncertain switched stochastic systems with time-delay. Both almost sure (sample) stability and stability in mean square are investigated. Based on Lyapunov functional methods and **linear matrix inequality** techniques, new criteria for exponential robust stability of switched stochastic delay systems with **non-linear uncertainties** are derived in terms of linear matrix inequalities and **average dwell-time** conditions. Numerical examples are also given to illustrate the results.

(a) Present Keyphrases
Target: {stochastic systems; non-linear uncertainties; exponential stability; linear matrix inequality; average dwell-time}
CopyRNN: 1. **linear matrix inequality**, 2. switched stochastic systems, 3. robust stability, 4. **exponential stability**, 5. **average dwell-time**
TG-Net: 1. **exponential stability**, 2. switched stochastic systems, 3. **average dwell-time**, 4. **non-linear uncertainties**, 5. **linear matrix inequality**

(b) Absent Keyphrases
Target: {switched systems; time-delay system}
CopyRNN: 1. **switched systems**, 2. switched delay systems, 3. robust control, 4. uncertain systems, 5. switched stochastic stochastic systems
TG-Net: 1. almost sure stability, 2. **switched systems**, 3. **time-delay systems**, 4. mean square stability, 5. uncertain systems

Figure 4: A prediction example of CopyRNN and TG-Net. The top 5 predictions are compared and the correct predictions are highlighted in bold.

	Present		Absent	
Model	F1@5	F1@10	R@10	R@50
TG-Net	**0.349**	**0.295**	**0.146**	**0.253**
-title	0.334	0.288	0.142	0.240
-copy	0.306	0.281	0.127	0.216

Table 5: Ablation study on **Krapivin** dataset.

As shown in Table 5, after we remove the title-guided part and only reserve the sequence encoding for the context (i.e., -title), both the present and absent keyphrase prediction performance become obviously worse, indicating that our title-guided context encoding is consistently critical for both present and absent keyphrase generation tasks. We also investigate the effect of removing the copy mechanism (i.e., -copy) from our TG-Net. From Table 5, we notice that the scores decrease dramatically on both present and absent keyphrase prediction, which demonstrates the effectiveness of the copy mechanism in finding important parts of the context.

Case Study

A keyphrase prediction example for a paper about the exponential stability of uncertain switched stochastic delay systems is shown in Figure 4. To be fair, we also only compare the RNN-based models (i.e., TG-Net and CopyRNN). For present keyphrase, we find that a present keyphrase “non-linear uncertainties”, which is a title phrase, is correctly predicted by our TG-Net, while CopyRNN fails to do so. As for absent keyphrase, we note that CopyRNN fails to predict the absent keyphrase “time-delay systems”. But our TG-Net can effectively utilize the title information “stochastic delay systems” to locate the important abstract information “stochastic systems with time-delay” and then successfully generate this absent keyphrase. These results exhibit that our model is capable of capturing the title-related core information more effectively and achieving better results in predicting present and absent keyphrases.

Conclusion

In this paper, we propose a novel TG-Net for keyphrase generation task, which explicitly considers the leading role of the title to the overall document main body. Instead of simply concatenating the title and the main body as the only source input, our model explicitly treats the title as an extra query-like input to guide the encoding of the context. The proposed TG-Net is able to sufficiently leverage the highly summative information in the title to guide keyphrase generation. The empirical experiment results on five popular real-world datasets exhibit the effectiveness of our model for both present and absent keyphrase generation, especially for a document with very low or very high title length ratio. One interesting future direction is to explore more appropriate evaluation metrics for the predicted keyphrases instead of only considering the exact match with the human labeled keyphrases as the current *recall* and *F-measure* do.

Acknowledgments

The work described in this paper was partially supported by the Research Grants Council of the Hong Kong Special Administrative Region, China (No. CUHK 14208815 and No. CUHK 14210717 of the General Research Fund), and Microsoft Research Asia (2018 Microsoft Research Asia Collaborative Research Award). We would like to thank Jingjing Li, Hou Pong Chan, Piji Li and Lidong Bing for their comments.

Earum vel eveniet quidem soluta, eligendi est sequi vitae blanditiis impedit dolores at cum culpa sunt consectetur, adipisci reiciendis obcaecati?Molestiae libero sequi ab dignissimos dicta quidem laudantium dolores, dicta quo unde culpa voluptatum id, rem minima incidunt fugit commodi non, itaque nostrum delectus, sunt placeat nostrum repellendus a harum quaerat dolore blanditiis?Aspernatur omnis quasi odit harum qui velit voluptatum odio nesciunt minima, ab harum reprehenderit quia nisi reiciendis soluta at?Distinctio ullam aliquid, autem nulla ullam ipsa inventore fugiat voluptas?Nulla voluptatum et quidem in nihil commodi earum dolor, molestias quibusdam libero nobis maxime excepturi doloribus beatae assumenda corrupti, repellat nemo sequi pariatur, esse aut in dignissimos, reiciendis vel provident.Tenetur soluta magnam sapiente dolores mollitia, incidunt totam nulla voluptate.Quibusdam officia fuga deserunt neque ducimus nulla libero veniam totam voluptas, quis quod quas nihil aperiam debitis excepturi, molestiae quam iure dicta alias amet.