Improving Convergence in IRGAN with PPO

Moksh Jain
Undergraduate, Department of Information Technology
NIT Karnataka, Surathkal, India

Sowmya Kamath S
Department of Information Technology
NIT Karnataka, Surathkal, India

ABSTRACT

Information retrieval modeling aims to optimise generative and discriminative retrieval strategies, where, generative retrieval focuses on predicting query-specific relevant documents and discriminative retrieval tries to predict relevancy given a query-document pair. IRGAN unifies the generative and discriminative retrieval approaches through a minimax game. However, training IRGAN is unstable and varies largely with the random initialization of parameters. In this work, we propose improvements to IRGAN training through a novel optimization objective based on proximal policy optimisation and gumbel-softmax based sampling for the generator, along with a modified training algorithm which performs the gradient update on both the models simultaneously for each training iteration. We benchmark our proposed approach against IRGAN on three different information retrieval tasks and present empirical evidence of improved convergence.

CCS CONCEPTS

• **Information systems → Novelty in information retrieval**; *Learning to rank*.

KEYWORDS

information retrieval, generative models, policy optimization

ACM Reference Format:
Moksh Jain and Sowmya Kamath S. 2020. Improving Convergence in IRGAN with PPO. In *7th ACM IKDD CoDS and 25th COMAD (CoDS COMAD 2020), January 5–7, 2020, Hyderabad, India.* ACM, New York, NY, USA, 2 pages. https://doi.org/10.1145/3371158.3371209

1 INTRODUCTION

IRGAN[8] unifies generative and discriminative retrieval models, in the framework of Generative Adversarial Networks[2], through a theoretical minimax game, allowing iterative optimization of both the models. The training of the generator in IRGAN can be formulated as a single-step reinforcement learning problem. Proximal Policy Optimization has achieved state-of-the-art performance in many reinforcement learning tasks[7]. The Gumbel-Softmax reparameterization trick[4] has been applied to a variety of problem domains with great success [5, 9]. We incorporate these ideas along with a modified training procedure to improve IRGAN performance.

CoDS COMAD 2020, January 5–7, 2020, Hyderabad, India

ACM ISBN 978-1-4503-7738-6/20/01…$15.00
https://doi.org/10.1145/3371158.3371209

2 METHODOLOGY

Consider the general information retrieval problem: given a query set $\{q_1 \dots q_N\}$, and a set of documents $\{d_1 \dots d_M\}$, select a subset of relevant documents for each query. IRGAN [8] consists of a minimax game between a generative model $p_\theta(d|q, r)$, and a discriminative model $f_\phi(q, d)$. The training of the generative model can be formulated as a single-step reinforcement learning problem. We propose a training objective based on proximal policy optimization [7] and the gumbel-softmax[4] based sampling of documents from the generative model as follows ($v_1 \dots v_M$ are i.i.d samples from Gumbel(0, 1) and τ is softmax temperature):

$$J^G(q_i) = \mathbf{E}_{d\sim p_{\theta'}(d|q_i,r)}[\min(r_i(\theta)A_i^{p_{\theta'}(d|q,r)}, clip(r_i(\theta), 1+\epsilon, 1-\epsilon)A_i^{p_{\theta'}(d|q,r)})]$$

$$r_i(\theta) = \frac{p_\theta(d|q,r)}{p_{\theta'}(d|q,r)}$$

$$A_i^{p_\theta(d|q,d)} = \log(1+\exp(f_\phi(q,d))) - \mathbf{E}_{d\sim p_{d|q,r}}[\log(1+\exp(f_\phi(q,d)))]$$

$$\mathrm{p}_\theta(d_i|q,r) = \frac{\exp((\log g_\theta(q,d_i)+v_i)/\tau)}{\sum_{k=1}^{M}\exp((\log g_\theta(q,d_k)+v_k)/\tau)}$$

The proposed training algorithm involves simultaneous updates to θ and ϕ for each iteration in training. We also maintain a target generator network with parameter θ' which is updated every k iterations to match the current value of θ.

3 RESULTS

Table 1 summarizes the observed results on LETOR 4.0 (MQ2008)[6] for web search, MovieLens-100k (ML-100k)[3] for item recommendation and InsuranceQA (IQA) [1] for question answering on standard metrics like Precision@k and Normalized Discounted Cumulative Gain@k using the same experimental setup as [8]. Our model achieved significant performance improvement of around 7.5 - 11% over IRGAN on all tested tasks. We observe improved performance across tasks. This increase in precision indicates that the generator learns a better estimate of the underlying relevance distribution, resulting in a higher fraction of relevant documents being retrieved. The increased *ndcg* scores indicate the improved graded relevance of the retrieved documents. These observations indicate that, our model's training converges to a strategy closer to the Nash equilibrium of the minimax game than in the standard IRGAN.

Table 1: Observed results for various IR tasks

Model	MQ2008		ML-100k		IQA
	p@5	ndcg@5	p@10	ndcg@10	p@1
IRGAN	0.1657	0.2225	0.3140	0.3723	0.6444
Our Model	**0.1860**	**0.2418**	**0.3378**	**0.4026**	**0.7165**

REFERENCES

[1] Minwei Feng, Bing Xiang, Michael R Glass, Lidan Wang, and Bowen Zhou. 2015. Applying deep learning to answer selection: A study and an open task. In *2015 IEEE Workshop on Automatic Speech Recognition and Understanding (ASRU)*. IEEE, 813–820.

[2] Ian Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio. 2014. Generative adversarial nets. In *Advances in neural information processing systems*. 2672–2680.

[3] F Maxwell Harper and Joseph A Konstan. 2016. The movielens datasets: History and context. *Acm transactions on interactive intelligent systems (tiis)* 5, 4 (2016), 19.

[4] Eric Jang, Shixiang Gu, and Ben Poole. 2016. Categorical reparameterization with gumbel-softmax. *arXiv preprint arXiv:1611.01144* (2016).

[5] Ryan Lowe, Yi Wu, Aviv Tamar, Jean Harb, OpenAI Pieter Abbeel, and Igor Mordatch. 2017. Multi-agent actor-critic for mixed cooperative-competitive environments. In *Advances in Neural Information Processing Systems*. 6379–6390.

[6] Tao Qin, Tie-Yan Liu, Jun Xu, and Hang Li. 2010. LETOR: A benchmark collection for research on learning to rank for information retrieval. *Information Retrieval* 13, 4 (2010), 346–374.

[7] John Schulman, Filip Wolski, Prafulla Dhariwal, Alec Radford, and Oleg Klimov. 2017. Proximal policy optimization algorithms. *arXiv preprint arXiv:1707.06347* (2017).

[8] Jun Wang, Lantao Yu, Weinan Zhang, Yu Gong, Yinghui Xu, Benyou Wang, Peng Zhang, and Dell Zhang. 2017. Irgan: A minimax game for unifying generative and discriminative information retrieval models. In *Proceedings of the 40th International ACM SIGIR conference on Research and Development in Information Retrieval*. ACM, 515–524.

[9] Zichao Yang, Zhiting Hu, Ruslan Salakhutdinov, and Taylor Berg-Kirkpatrick. 2017. Improved variational autoencoders for text modeling using dilated convolutions. In *Proceedings of the 34th International Conference on Machine Learning-Volume 70*. JMLR. org, 3881–3890.