

CERTIFICATE

It is certified that the work contained in the thesis titled Policy Gradient Reinforcement Learning for Ranking in Search and Recommender Systems by Vaibhav Padhye has been carried out under my/our supervision and that this work has not been submitted elsewhere for a degree.

It is further certified that the student has fulfilled all the requirements of Comprehensive Examination, Candidacy and SOTA for the award of Doctorate of Philosophy.

K. Lakshmanan

Supervisor

Prof. Kailasam Lakshmanan

Department of Computer Science

Indian Institute Of Technology(BHU)

Varanasi 221 005

DECLARATION BY THE CANDIDATE

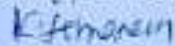
I, Vaibhav Padhye, certify that the work embodied in this thesis is my own bona fide work and carried out by me under the supervision of Prof. Kailasam Lakshmanan from December 2017 to December 2023, at the Department of Computer Science, Indian Institute of Technology, Varanasi. The matter embodied in this thesis has not been submitted for the award of any other degree/diploma. I declare that I have faithfully acknowledged and given credits to the research workers wherever their works have been cited in my work in this thesis. I further declare that I have not willfully copied any other's work, paragraphs, text, data, results, etc., reported in journals, books, magazines, reports dissertations, theses, etc., or available at websites and have not included them in this thesis and have not cited as my own work.

Date: 24-12-2024
Place: Varanasi, India


Signature of the Student

CERTIFICATE BY THE SUPERVISOR

It is certified that the above statement made by the student is correct to the best of my/our knowledge.


Supervisor

Prof. Kailasam Lakshmanan
Department of Computer Science
Indian Institute Of Technology(BHU)
Varanasi 221 005

पर्यवेक्षक/Supervisor
संगणक विज्ञान एवं अभियांत्रिकी विभाग
Department of Computer Sc & Engg
भारतीय प्रौद्योगिकी संस्थान
Indian Institute of Technology
(संगणक विज्ञान एवं अभियांत्रिकी विभाग,
सिन्दू विश्वविद्यालय,
Varanasi-221005)


Signature of Head of Department

आचार्य व विभागाध्यक्ष
Professor & Head
संगणक विज्ञान एवं अभियांत्रिकी विभाग
Department of Computer Sc. & Engg
भारतीय प्रौद्योगिकी संस्थान

COPYRIGHT TRANSFER CERTIFICATE

Title of the Thesis: Policy Gradient Reinforcement Learning for Ranking in Search and Recommender Systems

Name of the Student: Vaibhav Padhye

COPYRIGHT TRANSFER

The undersigned hereby assigns to the Indian Institute of Technology, Varanasi all rights under copyright that may exist in and for the above thesis submitted for the award of the Doctor of Philosophy.

Date: 24-12-2024
Place: Varanasi, India



Signature of the Student

Note: However, the author may reproduce or authorize others to reproduce material extracted verbatim from the thesis or derivative of the thesis for author's personal use provided that the source and the Institute's copyright notice are indicated.

ACKNOWLEDGEMENT

First of all, I would like to start by thanking my parents for always giving me support and motivation while carrying out thesis work. I would like to thank my supervisor Dr Kailasam Lakshmanan for his constant help monitoring and guidance throughout the duration of my thesis work. His support, inputs and guidance has been highly beneficial and of utmost importance while carrying out research work. I

am highly grateful to all the members of my Research Program Evaluation Committee, Dr. Ravindranath Choudhary and Dr. Sukhada for their valuable suggestions and guidance. Next i would like to thank my colleagues Mr. Ashwini Kumar, Ms. Sheetal Arya, Mr Saurabh Arora, Mr. Venkat Krishna and others in the department for their assistance. Last but not the least, I would thank God Almighty to provide me strength and patience to pursue this work over the years.

List of Figures

1.1	Reinforcement learning framework	4
1.2	Actor Critic framework	9
3.1	State Representation	42
3.2	Reward Distribution	53
3.3	Varying Discount factor	53
3.4	NDCG metric with different discount factor, x-axis: Discount factor, y-axis: NDCG	54
3.5	Varying Batch Size	54
3.6	Varying Episode Length	55
3.7	Comparing DRLRank with other DRL algorithms	55
4.1	MADDPG Architecture	63
4.2	State/Observation representaion module	65
4.3	Reward Distribution with varying Discount Factor on MSLR Web 30k	72
4.4	Reward Distribution with varying Discount Factor on MSLR Web 10k	73
4.5	Performance of the algorithm compared with the single agent ver- sion on MSLR-Wb 10k (left), MSLR-Web 30k (right), N denotes the number of agents	73
4.6	NDCG metric for different number of agents (a)MSLR -Web 10k (b) MSLR-Web30k	73
5.1	Flowchart depicting the modeling of recommendation process	89

5.2	Reward Distribution of MDP over 2 datasets	93
5.3	Parameter Sensitivity by varying Clipping ratio	94
5.4	Result of Varying Discount factor	94
5.5	Effect of different embedding sizes on Precision	95
5.6	Cold User Threshold	95

List of Tables

3.1	Statistics on OHSUMED, MQ2007, MQ2008 Datasets	51
3.2	Hyperparameters	51
3.3	Results for different metrics on OHSUMED Dataset	52
3.4	Results for different metrics on MQ2007 dataset	52
3.5	Results for different metrics on MQ2008 Dataset	52
4.1	Statistics for MSLR Web 10k and MSLR Web 30k Datasets	72
4.2	Hyperparameters	72
4.3	Results for different metrics on MSLR Web-10k Dataset	74
4.4	Results for different metrics on MSLR Web-30k Dataset	74
5.1	Hyperparameters	91
5.2	Results for different metrics on Movielens 100k	92
5.3	Results for different metrics on Movielens 1M	92
5.4	Wilcoxon Signed P value against different baselines on Movielens 100k	95
5.5	Wilcoxon Signed P value against different baselines on Movielens 1M	96