

EFFECTIVE INFERENCE & LEARNING IN STATISTICAL RELATIONAL MODELS

HAPPY MITTAL



**DEPARTMENT OF COMPUTER SCIENCE & ENGINEERING
INDIAN INSTITUTE OF TECHNOLOGY DELHI
NOVEMBER 2021**

© Indian Institute of Technology Delhi (IITD), New Delhi, 2021

EFFECTIVE INFERENCE & LEARNING IN STATISTICAL RELATIONAL MODELS

by

HAPPY MITTAL

Department of Computer Science & Engineering

Submitted

**in fulfilment of the requirements of the degree of Doctor of Philosophy
to the**



INDIAN INSTITUTE OF TECHNOLOGY DELHI

NOVEMBER 2021

Certificate

This is to certify that the thesis titled **Effective Inference & Learning in Statistical Relational Models** being submitted by **Mr. Happy Mittal** for the award of **Doctor of Philosophy in Computer Science and Engineering** is a record of bona fide work carried out by him under my guidance and supervision at the **Department of Computer Science and Engineering, Indian Institute of Technology Delhi**. The work presented in this thesis has not been submitted elsewhere, either in part or full, for the award of any other degree or diploma unless otherwise stated explicitly. In particular, work in Chapters 3, 5, and 6 was done jointly with undergraduate students. Work in some parts of Chapters 4 and 5 was also done jointly with master's students. In each case, the part done by the collaborators appeared in their respective bachelor's or master's theses.

Dr. Parag Singla
Associate Professor
Department of Computer Science and Engineering
Indian Institute of Technology Delhi
New Delhi- 110016

Acknowledgements

I wish to thank my advisor Dr. Parag Singla, who guided me in this journey of Ph.D. His command over the subject helped me get started very early in my Ph.D. career. He taught me how to think clearly about the problem, and how to write down ideas floating in mind to bring out a research paper. I remember him saying in my first year that I will earn my Ph.D. only when I will be able to write a full research paper by myself. I think this is a very good benchmark for anyone who wants to analyze himself/herself whether he/she is ready to finish a Ph.D. Moreover, Parag is a great human being. I have learned a lot of traits from him and I wish to follow his footsteps rest of my life also. I would also like to thank Dr. Mausam, who at various times, gave honest and critical feedback on my projects. He also helped me decide the career path after the Ph.D. and presented both pros and cons of the options I had. I am also grateful to Dr. Vibhav Gogate, with whom we collaborated for all our research projects, and that collaboration turned out to be great. He has a great knowledge of his subject, which was reflected in the various meetings I have had with him. I also thank Somdeb Sarkhel with whom I exchanged many research ideas over the emails, and he always helped me with his vast knowledge in the area of SRL.

I am grateful to my parents and my brother Nicky Mittal who supported me throughout this journey. I also want to thank my wife Priyanka who constantly supported me during my endeavour. Moreover, the kids (Akshita and Samaira) in my family bugged me a lot during my thesis writing. Finally, a special mention to my little one Aradhya because of whom I had to spend many sleepless nights.

During my PhD life, I developed a great friendship with my Ph.D. colleagues: Ankit Anand, Prachi Jain, Vishal Sharma, Yashoteja Prabhu, Himanshu Jain, Dilpreet Kaur, Dinesh Khandelwal, and many more with whom I had both technical and non-technical discussions, helping me reduce the stress of academic life.

I worked with many bachelor's and master's students during my research, and because of their support and hard work, I was able to publish good quality work. I would like to thank Prasoon Goyal, Anuj Mahajan, Shubhankar Singh, Noman Sheikh, Piyush Gupta, and Ayush Bhardwaj who spent numerous sleepless nights working and experimenting in labs.

I would also like to thank Prof. Kristian Kersting who hosted me for a 3 months research internship at Dortmund University. I learned a lot from him, and also got a flavor of how research is conducted outside Indian universities.

Finally, I am thankful to Tata Consultancy Services for funding my Ph.D. during the past 4 years that helped me attending and presenting my work at various international conferences.

Happy Mittal

Abstract

A vast majority of traditional Machine Learning algorithms make the i.i.d. assumption, i.e., instances are assumed to be sampled independently from an identical distribution. However, the real world data is relational i.e. there exist various types of relations among the instances which need to be captured. Moreover, these relations also have a certain degree of uncertainty that needs to be modeled. The field of Statistical Relational Learning (SRL) achieves the merger between the two by explicitly modeling relations as well as uncertainty among them. Due to the underlying template structure, SRL models often result in the construction of large ground networks posing serious computational challenges. Therefore, designing effective inference and learning algorithms becomes critical for the success of these models. This thesis focuses on proposing multiple such inference and learning techniques: (a) We first propose two different algorithms for lifted inference, i.e., techniques which exploit the symmetry of the underlying relational representation. In the first approach, we exploit the underlying symmetry using a rule called Single Occurrence (SO) for MAP inference. SO rule results in inference time which is independent of domain size under certain conditions. Second, we propose a language for specifying constraints

for constructing the underlying ground theory and extend existing lifting techniques to work with these constraints.

(b) In the second part of this thesis, we present two different techniques for effective learning in SRL models. First, we present an algorithm for fine-grained learning, where we learn a different weight for different subsets of the groundings of the same logical formula. We show that this can result in improved prediction accuracy by reducing bias. Second, we present a novel re-parameterization technique to generalize the weights learned for logical formulas across domains of varying sizes. We also present various theoretical properties of our formalism.

In each of the above cases, we present experimental results validating the efficacy of proposed algorithms on a mix of real and artificial datasets. Much of our work is presented in the context of Markov Logic Networks (MLNs), which is a popular SRL formalism though our techniques are general enough and can be extended to other similar representations. We conclude the thesis with directions for future work.

सार

पारंपरिक मशीन लर्निंग एल्गोरिथ्म का एक विशाल बहुमत i.i.d. धारणा, यानी, उदाहरणों को एक समान एवं स्वतंत्र रूप से वितरित मानता है। यद्यपि वास्तविक विश्व डेटा संबंधपरक है अर्थात् उदाहरणों के बीच विभिन्न प्रकार के संबंध मौजूद हैं जिन्हें पकड़ने की आवश्यकता है। इसके अतिरिक्त, इन संबंधों में कुछ कुछ मात्रा में अनिश्चितता भी होती है जिसे प्रतिरूपित करने की आवश्यकता होती है। स्टैटिस्टिकल रिलेशनल लर्निंग (एसआरएल) का क्षेत्र स्पष्ट रूप से मॉडलिंग संबंधों के साथ-साथ उनके बीच अनिश्चितता के बीच विलय को प्राप्त करता है। अंतर्निहित टेम्पलेट संरचना के कारण, एसआरएल मॉडल अक्सर बड़े ग्राउंड नेटवर्क का निर्माण करते हैं जो गंभीर कम्प्यूटेशनल चुनौतियों का सामना करते हैं। इसलिए, इन मॉडलों की सफलता के लिए प्रभावी इन्फेरेंस और लर्निंग के एल्गोरिथ्म को डिजाइन करना महत्वपूर्ण हो जाता है। यह थीसिस कई ऐसे इन्फेरेंस और लर्निंग की तकनीकों को प्रस्तावित करने पर केंद्रित है: (ए) हम पहले लिफ्टेड इन्फेरेंस के लिए दो अलग-अलग एल्गोरिथ्म का प्रस्ताव करते हैं, यानी ऐसी तकनीकें जो अंतर्निहित संबंधपरक निरूपण की समरूपता का फायदा उठाती हैं। पहले दृष्टिकोण में, हम एमएपी अनुमान के लिए सिंगल अकरैस (एसओ) नामक नियम का उपयोग करके अंतर्निहित समरूपता का फायदा उठाते हैं। SO नियम का परिणाम इन्फेरेंस समय होता है जो कुछ शर्तों के तहत डोमेन आकार से स्वतंत्र होता है। दूसरा, हम अंतर्निहित ग्राउंड थ्योरी के निर्माण के लिए बाधाओं को निर्दिष्ट करने के लिए एक भाषा का प्रस्ताव करते हैं और इन बाधाओं के साथ काम करने के लिए मौजूदा लिफ्टिंग की तकनीकों का विस्तार करते हैं। (बी) इस थीसिस के दूसरे भाग में, हम एसआरएल मॉडल में प्रभावी लर्निंग के लिए दो अलग-अलग तकनीकों को प्रस्तुत करते हैं। सबसे पहले, हम अच्छी तरह से लर्निंग के लिए एक एल्गोरिथ्म प्रस्तुत करते हैं, जहां हम एक ही तार्किक सूत्र के आधार के विभिन्न उपसमुच्चय के लिए एक अलग वजन सीखते हैं। हम दिखाते हैं कि इससे पूर्वाग्रह को कम करके भविष्यवाणी की सटीकता में सुधार हो सकता है। दूसरा, हम अलग-अलग आकार के डोमेन में तार्किक सूत्रों के लिए सीखे गए भार को सामान्य बनाने के लिए एक नया पुनः पैरामीट्रीकरण तकनीक प्रस्तुत करते हैं। हम अपने सूत्रवाद के विभिन्न सैद्धांतिक गुण भी प्रस्तुत करते हैं। उपरोक्त प्रत्येक मामले में, हम वास्तविक और साथ ही कृत्रिम डेटासेट पर प्रस्तावित एल्गोरिथ्म की प्रभावकारिता को मान्य करते हुए प्रयोगात्मक परिणाम प्रस्तुत करते हैं। हमारा अधिकांश काम मार्कोव लॉजिक नेटवर्क्स (एमएलएन) के संदर्भ में प्रस्तुत किया गया है, जो एक लोकप्रिय एसआरएल सूत्रवाद है, हालांकि हमारी तकनीक काफी सामान्य है और इसे अन्य समान निरूपणों तक बढ़ाया जा सकता है।

Contents

Certificate	i
Acknowledgements	iii
Abstract	v
List of Figures	xvii
List of Tables	xix
I Prologue	1
1 Introduction	3
1.1 Overview	3

1.2	Statistical Relational Learning	5
1.3	Advances in Learning and Inference	11
1.3.1	Lifted Inference	12
1.3.2	Effective Learning	15
1.4	Contributions	17
2	Background	21
2.1	First Order Logic	21
2.2	Markov Logic Networks	22
2.3	Inference in Markov Logic networks	24
2.3.1	Inference tasks	24
2.3.2	Inference Methods	26
2.3.3	Lifted Inference	26
2.4	Lifted Inference Rules	28
2.4.1	Decomposer Rule	28
2.4.2	Binomial Rule	29
2.5	Learning Weights in Markov Logic Networks	30

II	Effective Inference	33
3	New Rules For Domain Independent Lifted MAP Inference	35
3.1	Introduction	35
3.2	Basic Framework	37
3.2.1	Motivation	37
3.2.2	Notation and Preliminaries	38
3.3	Exploiting Single Occurrence	39
3.3.1	First Rule for Lifting MAP	39
3.3.2	Relationship to Existing Rules	43
3.3.3	Domain Independent Lifted MAP	44
3.4	Exploiting Extremes	46
3.4.1	Extreme Assignments	46
3.4.2	Second Rule for Lifting MAP	47
3.4.3	A Procedure for Identifying Tautologies	49
3.5	Experiments	50
3.5.1	Datasets and Methodology	50

3.5.2	Results	51
3.6	Conclusion & Future work	53
4	Lifted Inference Rules With Constraints	55
4.1	Introduction	55
4.2	Constraint Language	57
4.2.1	Language Specification	58
4.2.2	Canonical Representation	59
4.2.3	Operators in the Language	64
4.3	Extending Lifted Inference Rules	68
4.3.1	Motivation & Key Operations	68
4.3.2	Decomposer	72
4.3.3	Binomial	74
4.3.4	Single Occurrence	80
4.3.5	Normal Forms and Evidence Processing	82
4.4	Experiments	83
4.4.1	Exact Inference	84

4.4.2	Approximate Inference	85
4.5	Conclusion & Future Work	87

III Effective Learning 89

5 Fine-Grained Weight Learning in Markov Logic Networks 91

5.1	Introduction	91
5.2	Related Work	94
5.3	Introducing Subtypes In MLN	94
5.4	Learning Weights	96
5.4.1	Pipelined Approach	97
5.4.2	Joint Learning of SubTypes	99
5.5	Inference	101
5.6	Experiments	102
5.6.1	IMDB	102
5.6.2	WebKB	106
5.7	Conclusion & Future work	108

6	Domain-Size Aware Markov Logic Networks	111
6.1	Introduction	111
6.2	Related Work	113
6.3	Characterizing Issues in MLNs	115
6.3.1	Motivation	115
6.3.2	Earlier work	116
6.3.3	A More Correct Analysis	118
6.4	Domain-size Aware Markov Logic Networks	127
6.4.1	Intuition	127
6.4.2	Definitions	127
6.4.3	Inference and Learning	130
6.4.4	Properties	130
6.5	Handling Evidence	135
6.6	Experiments	136
6.6.1	Datasets and Methodology	137
6.6.2	Results	138

6.7 Conclusion and Future Work	139
--	-----

IV Epilogue	141
--------------------	------------

7 Conclusion & Future Work	143
---------------------------------------	------------

7.1 Conclusion	143
--------------------------	-----

7.2 Future Work	144
---------------------------	-----

Bibliography	147
---------------------	------------

Publications Resulting From The Thesis	161
---	------------

Biography	163
------------------	------------

List of Figures

1.1	Markov Network for the social networking problem. Here <i>Ana</i> is denoted by <i>A</i> and <i>Bob</i> by <i>B</i>	8
1.2	General framework of lifted inference	12
2.1	Markov Network for the MLN <i>M</i> given in Example 1.	24
3.1	Splitting an MLN theory	42
3.2	IE	53
3.3	Friends & Smokers	53
3.4	student	54
4.1	Results for exact inference on FS	86
4.2	Results for exact inference on PS	86

4.3	Results using approximate inference on WebKB and IMDB	87
5.1	Parameter learning of an MLN by three approaches. Our approach finds the right balance between bias and variance	93
5.2	AUC on IMDB dataset for different number of cluster (with varying evi- dence)	105
5.3	AUC comparison on IMDB dataset with pipelined vs joint learning with varying amount of evidence($k=2$)	105
5.4	AUC comparison on WebKB dataset with varying number of clusters . . .	108
6.1	Figure showing effect of increasing domain size of x (denoted by $ \Delta_x $) on $Pr(\text{epidemic})$ in MLN formula $w : \text{sick}(x) \Rightarrow \text{epidemic}$. Let $w = 1$. Figures 6.1a and 6.1b show the Ground Network and $Pr(\text{epidemic})$ for $ \Delta_x = 3$ and $ \Delta_x = 15$ respectively. Figure 6.1c shows the plot of $ \Delta_x $ vs $Pr(\text{epidemic})$ in MLNs and DA-MLNs (our approach). For readability, sick and epidemic are abbreviated as S and S respectively.	113
6.2	Query Marginal vs Weight (Proposition 6.4.3)	133
6.3	Query Marginal vs Weight (Proposition 6.4.4)	134
6.4	Results and Rules for FS, IMDB, and WebKB datasets.	140

List of Tables

2.1	Categorization of lifted inference techniques	27
4.1	A comparison of constraint languages proposed in literature across four dimensions. The deficiencies/missing properties for each language have been highlighted in bold. Among the existing work, only KC allows for a full set of constraints. GCFOVE (tree-based) and LBP (hypercubes) allow for subset constraints but they do not explicitly handle inequality. PTP does not handle subset constraints. For constraint aggregation, most approaches allow only intersection of atomic constraints. GCFOVE and LBP allow union of intersections (DNF) but only deal with subset constraints. See footnote 4 in Broeck [32] regarding KC. Lifted VE, KC and PTP use a general purpose constraint solver which may not be tractable. Our approach allows for all the features discussed above and uses a tractable solver. We propose a constrained solution for lifted search and sampling. Among earlier work, only PTP has looked at this problem (both search and sampling). However, it only allows a very restrictive set of constraints.	56

4.2	Split of subset constraint over x_1	61
4.3	Dataset Details. var : domain size varied. '+' : a separate weight learned for each grounding	84
5.1	Rules of IMDB MLN	103
5.2	Rules of IMDB MLN with subtype predicate	104
5.3	Rules of WebKB MLN	106
5.4	Rules of modified WebKB MLN	107