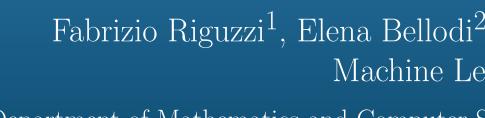
PROBABILISTIC INDUCTIVE CONSTRAINT LOGIC



Probabilistic Constraint Logic Theories

Inductive Constraint Logic (De Raedt and Van Laer): learning Constraint Logic Theories (sets of integrity constraints) from interpretat rather than from entailment

In this work: Probabilistic Constraint Logic Theories (PCLTs) and a system (PASCAL) to perform discriminative learning of their structure and parameters from interpretations.

Language A PCLT is a set of probabilistic integrity constraints (ICs) of the form $p_i :: L_1, \ldots, L_b \to \exists (P_1); \ldots; \exists (P_n); \forall \neg (N_1); \ldots; \forall \neg (N_m)$

Semantics A PCLT T defines a probability distribution on the set W of ground constraint logic theories called *possible theories*; for each grounding of the body of each IC, we include the IC in a possible theory with probability p_i and we assume all groundings to be independent.

An IC's probability is the sum of the probabilities of the possible theories where a grounding of the constraint is present.

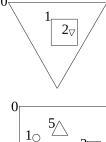
In each possible theory, An IC C is true in an interpretation I $(I \models C)$ if and only if, for each substitution θ such that each literal in $Body(C)\theta$ is ground and true in I, at least one disjunct in $Head(C)\theta$ is true in I.

The probability $P(\oplus|I)$ of the positive class given an interpretation I, the probability of a PCLT T satisfying I, is

$$P(\oplus|I) = \sum_{w \in W} P(\oplus, w|I) = \sum_{w \in W} P(\oplus|w, I) P(w|I) = \sum_{w \in W, M(\mathbf{BG} \cup I) \models w} P(w)$$

Example (from Bongard) Background knowledge: $in(A, B) \leftarrow inside(A, B)$ $in(A, D) \leftarrow inside(A, C), in(C, D)$ PCLT:

 $\{C_1 = 0.5 :: triangle(T), square(S), in(T, S) \rightarrow falls$



The body of C_1 is true for the sing substitution T/2 and S/1 thus m_1 1 and $P(\oplus|I) = 0.5$.

The body of C_1 is true for three pair (triangle, square) thus $m_1 = 3$ and $P(\oplus|I) = 0.125.$

PASCAL: PCLT structure learning

Problem Given

LE UN/1

- a set $\mathcal{I}^+ = \{I_1, \ldots, I_Q\}$ of positive interpretations
- a set $\mathcal{I}^- = \{I_{Q+1}, \ldots, I_R\}$ of negative interpretations
- a normal logic program **BG** (background knowledge)
- a language bias

find a PCLT T that maximizes the likelihood

$$L = \prod_{q=1}^{Q} P(\oplus|I_q) \prod_{r=Q+1}^{R} P(\ominus|I_r)$$

PASCAL solves this problem by first identifying good candidate ICs and then searching for a theory guided by the log likelihood (LL) of the data. Parameters:

- *MLB*, the maximum number of literals in the body of ICs;
- *MD*, the maximum number of disjuncts in the head of ICs;
- *MLP* and *MLN*, the maximum number of literals allowed in a P disjunct and a N disjunct respectively.

1: function PASCAL($\mathcal{I}^+, \mathcal{I}^-, \mathbf{BG}, NC, MLB, MD, MLP, MLN, BeamSize, MaxSteps$ Steps = 12:

2.	b c c p b = 1	
3:	$Beam \leftarrow (false \leftarrow true, -\infty)$	⊳ Empty
4:	repeat	
5:	NewBeam = []	
6:	while <i>Beam</i> is not empty do	\triangleright ICs sea
7:	Remove the first IC (C, LL) from <i>Beam</i>	
8:	$Ref \leftarrow all refinements of C respecting MLB, MD, MLP,$	and MLN
9:	for all $C' \in Ref \operatorname{\mathbf{do}}$	
10:	$(\{C''\}, LL'') \leftarrow \text{LearnParams}(\{C'\}, \mathcal{I}^+, \mathcal{I}^-, \mathbf{BG})$	\triangleright gradient des
11:	$NewBeam \leftarrow \text{INSERT}((C'', LL''), NewBeam)$	
12:	$\mathbf{if} \ size(NewBeam) > BeamSize \ \mathbf{then}$	
13:	Remove the last element of $NewBeam$	
14:	end if	
15:	end for	
16:	end while	
17:	$Beam \leftarrow NewBeam$	
18:	Steps = Steps + 1	
19:	$\mathbf{until}\ Steps > MaxSteps$	
20:	$T \leftarrow \emptyset, \ LL \leftarrow -\infty$	\triangleright Theory sea
21:	repeat	
22:	Remove the first couple (C, LL) from <i>Beam</i>	
23:	$(T', LL') \leftarrow \text{LearnParams}(T \cup \{C\}, \mathcal{I}^+, \mathcal{I}^-, \mathbf{BG})$	
24:	$\mathbf{if} \ LL' > LL \ \mathbf{then}$	
25:	$T \leftarrow T', LL \leftarrow LL'$	
26:	end if	
27:	until Beam is empty or T contains NC ICs	
28:	return T	
29:	end function	

Fabrizio Riguzzi¹, Elena Bellodi², Riccardo Zese², Marco Alberti¹, Evelina Lamma² Machine Learning 110(4): 723-754 (2021)

¹Department of Mathematics and Computer Science and ²Department of Engineering, University of Ferrara, Italy

						PC	CLT p	arameter	learning			
tions												
cture	1	Problem Give • a PCLT th										
		• a set $\mathcal{I}^+ =$	$= \{I_1, \ldots, I_n\}$	Q of position	tive interp	retations						
		• a set $\mathcal{I}^- =$	= $\{I_{Q+1}, \ldots$	$, I_R \}$ of ne	egative int	erpretati	ons					
	f	• a normal l ind the param		·	-		lge)					
se}		-					$L = \prod_{q=1}^{Q}$	$\left[P(\oplus I_q)\prod_{r=Q+1}^R\right]$	$P(\ominus I_r)$			
le =	<u>_</u>	The equation $\frac{\dot{c}}{\dot{c}}$	$\frac{\partial L}{\partial p_i} = 0$ does	s not admi	it a closed	form sol	ty that t ution, sc	he example lab we must use o eights are upda	els are observe ptimization to	find the ma	—	of L.
rs Id		$\mathbf{p}_{n+1} = \mathbf{p}_n - \epsilon \nabla_{\mathbf{p}} L(\mathbf{p}) = \mathbf{p}_n - \epsilon \nabla_{\mathbf{p}} \frac{\partial L}{\partial \mathbf{p}}$										
		parameters p_i ,	or with a s	second ord	er method	e of the l such as	step dor Limited-	e by gradient memory BFGS in most cases	descent along 6 (L-BFGS).	-	_	
]	Expe	rimental	results			
)		We compared 1 probabilistic)			0	hms LIF	TCOVE	R, SLIPCOVE	R and LEMU	R, the MLN	Is algorith	nms MLN-B
ty IC				verage AU						Ave	erage AU	C-ROC
earch		Dataset	SLIPCOVER	LIFT-EM	LIFT-LBFGS	6 PASCAL	TILDE		Dataset	SLIPCOVER	LIFT-EM	LIFT-LBFGS
		Bupa	1	1	1	1	0.420		Bupa Carcinogenesis	<u> </u>	$\frac{1}{0.766}$	<u> </u>
		Carcinogenesis Financial	0.745 0.173	0.672	$\frac{0.561}{0.187}$	$\frac{0.770}{0.317}$	$\begin{array}{r} 0.707 \\ \hline 0.123 \end{array}$		Financial	0.568	0.432	0.472
		Mondial	0.175	0.120	0.187	0.652	0.123 0.650		Mondial	0.630	0.663	0.643
escent		Mutagen.	0.920	0.971	0.725	0.902	0.851		Mutagen.	0.826	0.931	0.649
		Pyrimidine	0.956	1	0.819	0.990	0.769		Pyrimidine	0.925	1	0.850
		Sisya	0.708	0.706	0.706	0.622	0.621		Sisya Sisyb	0.719 0.500	0.372 0.500	0.721 0.500
		Sisvb	0.287			0.286			L Macardo			

v search



				_	ыхре	results						
PASCA	AL wi	th the P	LP algorith	nms LIF	TCOVE	ER and LEM	UR, the N	ALNs al	gorith	ms MLN-B	BC/MLN	J-BT, tł
relatio									-			
	Ave	rage AU	C-PR					Average	e AUC	-ROC		
SLIPCC	OVER	LIFT-EM	LIFT-LBFGS	PASCAL	TILDE	Dataset	SLIPCO	VER LIF	T-EM	LIFT-LBFGS	PASCAL	TILDE
						Bupa	1		1	1	1	0.500
		0.672	0.561	0.770	0.707	Carcinogene	esis 0.695	6 0.	766	0.472	0.763	0.667
						Financial	0.568	3 0.	432	0.535	0.745	0.478
		0.763			0.650	Mondial	0.630) 0.	663	0.643	0.495	0.500
0.92	20	0.971		0.902	0.851	Mutagen.	0.826	б О.	931	0.649	0.806	0.778
		1		0.990		Pyrimidine	0.925)	1	0.850	0.993	0.815
		0.706	0.706	0.622	0.621	Sisya	0.719) 0.	.372	0.721	0.502	0.499
0.28	87	0.286	0.286	0.286	0.286	Sisyb	0.500) 0.	500	0.500	0.500	0.500
0.560 0.734		0.734	0.760	0.855	0.685	Triazine	0.544 0.7		713	0.760	0.803	0.600
0.428 0.50		0.502	0.448	0.469	0.588	Yeast	0.733 0.786		786	0.721	0.794	0.718
0.89)9	0.966	0.970	0.635	0.300	Bongard	0.944	£ 0.	.975	0.987	0.749	0.500
		MINE	20	MINDT					MLN-B	C	MLN-BT	
LEMUR	MLN-F				PASCAL	Dataset	LEMUR	MLN-BC	samp.		samp.	PASCAL
		1		<u>^</u>		Carcinogenesi	is 0.691	0.619	0.633	0.503	0.494	0.770
						Mondial	0.864	0.585	0.742	0.735	0.781	0.652
0.952	0.690			_	0.902	Mutagenesis	0.952	0.690	0.831	0.872	_	0.902
	relation SLIPCO 1 0.74 0.77 0.92 0.95 0.76 0.28 0.56 0.42 0.89	relational cl Ave SLIPCOVER 1 0.745 0.173 0.776 0.920 0.956 0.956 0.708 0.287 0.560 0.428 0.560 0.428 0.899	Image: classifier of Average AU Average AU SLIPCOVER LIFT-EM 1 I 0.745 0.672 0.173 0.126 0.776 0.971 0.920 0.971 0.920 0.971 0.920 0.971 0.920 0.971 0.920 0.971 0.920 0.971 0.920 0.971 0.920 0.971 0.920 0.971 0.920 0.971 0.920 0.971 0.920 0.971 0.920 0.971 0.920 0.971 0.920 0.971 0.920 0.971 0.920 0.971 0.920 0.970 0.920 0.970 0.920 0.970 0.920 0.970 0.920 0.970 0.920 0.970 0.920 0.970 0.920 0.970 0.920 0.970 0.920	relational classifier TILDE. Average AUC-PR SLIPCOVER LIFT-EM LIFT-LBFGS 1 1 0.745 0.672 0.561 0.173 0.126 0.187 0.776 0.763 0.723 0.920 0.971 0.725 0.956 1 0.819 0.708 0.706 0.706 0.560 0.734 0.760 0.428 0.502 0.448 0.899 0.966 0.970 LEMUR MLN-BC samp. MLN-BT 0.691 0.619 0.633 0.503	PASCAL with the PLP algorithms LIF relational classifier TILDE. Average AUC-PR SLIPCOVER LIFT-EM LIFT-LBFGS PASCAL 1 1 1 0.745 0.672 0.561 0.770 0.173 0.126 0.187 0.317 0.776 0.763 0.723 0.652 0.920 0.971 0.725 0.902 0.956 1 0.819 0.990 0.708 0.706 0.706 0.622 0.287 0.286 0.286 0.286 0.428 0.502 0.448 0.469 0.899 0.966 0.970 0.635 MLN-BC MLN-BT MLN-BC MLN-BT MLN-BC MLN-BT 0.633 0.503 0.494	PASCAL with the PLP algorithms LIFTCOVER, SLIPCOV relational classifier TILDE. Average AUC-PR SLIPCOVER LIFT-EM LIFT-LBFGS PASCAL TILDE 1 1 1 0.420 0.745 0.672 0.561 0.770 0.707 0.173 0.126 0.187 0.317 0.123 0.776 0.763 0.723 0.652 0.650 0.920 0.971 0.725 0.902 0.851 0.956 1 0.819 0.990 0.769 0.708 0.706 0.706 0.622 0.621 0.287 0.286 0.286 0.286 0.286 0.428 0.502 0.448 0.469 0.588 0.899 0.966 0.970 0.635 0.300 MLN-BC MLN-BT MLN-BT samp. PASCAL 0.691 0.619 0.633 0.503 0.494 0.770	TILDE. Average AUC-PR SLIPCOVER LIFT-EM LIFT-LBFGS PASCAL TILDE 1 1 1 0.420 0.745 0.672 0.561 0.770 0.707 0.173 0.126 0.187 0.317 0.123 0.776 0.763 0.723 0.652 0.650 0.920 0.971 0.725 0.902 0.851 0.956 1 0.819 0.990 0.769 0.708 0.706 0.706 0.622 0.621 Sisya 0.560 0.734 0.760 0.855 0.685 0.286 0.286 0.286 0.560 0.734 0.760 0.855 0.685 0.685 0.286 0.428 0.502 0.448 0.469 0.588 0.899 0.966 0.970 0.635 0.300 MLN-BC MLN-BT samp. PASCAL Dataset Carcinogenesi MLN-BT	$\begin{array}{c c c c c c c c c c c c c c c c c c c $	$\begin{array}{c} \label{eq:particular} PASCAL with the PLP algorithms LIFTCOVER, SLIPCOVER and LEMUR, the MLNs all relational classifier TILDE. Average AUC-PR & Average \\ \hline SLIPCOVER LIFT-EM LIFT-LBFGS PASCAL TILDE \\ \hline 1 & 1 & 1 & 0.420 \\ \hline 0.745 & 0.672 & 0.561 & 0.770 & 0.707 \\ \hline 0.173 & 0.126 & 0.187 & 0.317 & 0.123 \\ \hline 0.0776 & 0.763 & 0.723 & 0.652 & 0.650 \\ \hline 0.956 & 1 & 0.819 & 0.990 & 0.769 \\ \hline 0.0287 & 0.286 & 0.286 & 0.286 & 0.286 \\ \hline 0.560 & 0.734 & 0.760 & 0.855 & 0.685 \\ \hline 0.428 & 0.502 & 0.448 & 0.469 & 0.588 \\ \hline 0.899 & 0.966 & 0.970 & 0.635 & 0.300 \\ \hline \\$	$\begin{array}{c} \label{eq:product} PASCAL with the PLP algorithms LIFTCOVER, SLIPCOVER and LEMUR, the MLNs algorith relational classifier TILDE. Average AUC-PR \\ \hline \\ \hline 1 & 1 & 1 & 1 & 0.420 \\\hline \hline 1 & 1 & 1 & 0.420 \\\hline \hline 0.745 & 0.672 & 0.561 & 0.770 & 0.707 \\\hline 0.776 & 0.763 & 0.723 & 0.652 & 0.650 \\\hline 0.920 & 0.971 & 0.725 & 0.902 & 0.851 \\\hline 0.956 & 1 & 0.819 & 0.990 & 0.769 \\\hline 0.708 & 0.706 & 0.706 & 0.622 & 0.621 \\\hline 0.287 & 0.286 & 0.286 & 0.286 & 0.286 \\\hline 0.560 & 0.734 & 0.760 & 0.635 & 0.668 \\\hline 0.939 & 0.966 & 0.970 & 0.635 & 0.300 \\\hline \\ \hline \\ $	Dataset SLIPCOVER LIFT-EM LIFT-LBFGS PASCAL TILDE. Average AUC-PR Average AUC-ROC SLIPCOVER LIFT-EM LIFT-LBFGS PASCAL TILDE 0.745 0.672 0.561 0.770 0.707 0.173 0.126 0.187 0.317 0.123 0.920 0.971 0.725 0.902 0.851 0.920 0.971 0.725 0.902 0.851 0.920 0.971 0.725 0.902 0.851 0.926 1 0.819 0.990 0.769 0.566 1 0.819 0.990 0.769 0.560 0.734 0.760 0.826 0.286 0.428 0.502 0.448 0.469 0.588 0.428 0.502 0.433 0.760 0.500 0.428 0.502 0.448 0.469 0.588 0.899 0.966 0.970 0.635 0.300 EMUR MLN-BC samp. PASCAL 0.691 0.61	PASCAL with the PLP algorithms LIFTCOVER, SLIPCOVER and LEMUR, the MLNs algorithms MLN-BC/MLN relational classifier TILDE. Average AUC-PR Average AUC-PR SLIPCOVER LIFT-EM LIFT-LBFGS PASCAL TILDE 1 1 1 0.420 0.745 0.672 0.561 0.770 0.707 0.173 0.126 0.187 0.317 0.123 0.920 0.971 0.725 0.902 0.851 0.926 1 0.819 0.990 0.769 0.708 0.706 0.763 0.723 0.652 0.685 0.926 1 0.819 0.990 0.769 0.560 0.774 0.760 0.632 0.631 0.449 0.806 0.926 0.734 0.760 0.685 0.685 0.786 0.721 0.792 0.926 1 0.819 0.990 0.769 0.776 0.500 0.500 0.500 0.500 0.926 0.734 0.760 0.885 0.886 0.286 0.286 0.286 0.993 0.500 0.500 0.500 0.50