LogicMP: A Neuro-symbolic Approach for Encoding First-order Logic Constraints

Weidi Xu[†‡], Jingwei Wang[‡], Lele Xie[‡], Jianshan He[‡], Hongting Zhou[‡], Taifeng Wang[◇], Xiaopei Wan[‡], Jingdong Chen[‡], Chao Qu[†], Wei Chu[‡†]

† INFLY TECH (Shanghai) Co., Ltd. ‡ Ant Group ◇ BioMap Research
wdxu@inftech.ai

Background

Neuro-symbolic Approachs

- Neural networks (NNs) are effective for representation learning.
- However, NNs are not necessary to obey the logical constraints.
- Neuro-symbolic methods aim to combine NNs with explicit logic.

An Example of Encoding First-order Logic Constraints.

- Task: Given the input image and input tokens, the task is to develop a function to predict whether two tokens coexist in a block.
- Rule: If tokens i and j are in the same block and tokens j and k are also together, then tokens i and k should be in the same block.

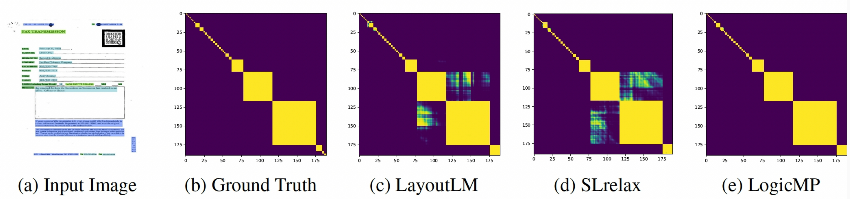

图 1: An example of using LogicMP in the image segmentation problem.

Markov Logic Networks

- Entities: the constants, e.g., two tokens e_1 and e_2.
- Predicates: the property or the relation, e.g., coexist predicate $\mathtt{C}$.
- Ground atom: the predicate with particular entities, e.g., $\mathtt{C}(e_1, e_2)$.
- Formula: e.g., $\forall a, b, c : \mathtt{C}(a, b) \wedge \mathtt{C}(b, c) \implies \mathtt{C}(a, c)$.
- Grounding: e.g., $\mathtt{C}(e_1, e_2) \wedge \mathtt{C}(e_2, e_3) \vee \mathtt{C}(e_1, e_3)$.

- Markov logic network (MLN) is an elegant probabilistic modeling with first-order logic, using the first-order logic as the joint potential.

$$p(\mathbf{v}|O) \propto \exp\Big(\underbrace{\sum_i \phi_u(v_i)}_{neural\ semantics} + \underbrace{\sum_{f\in F} w_f \sum_{g\in G_f} \phi_f(\mathbf{v}_g)}_{symbolic\ FOLCs} \Big)\,, \tag{1}$$

  - $\mathbf{v}/O$ is the set of unobserved/observed variables
  - neural semantics:
    - $\phi_u(\cdot) : v_i \mapsto \mathcal{R}$ models the evidence of single ground atom i in status v_i.
  - symbolic FOLCs:
    - w_f presents the weight of formula f
    - $\phi_f(\cdot) : \mathbf{v}_g \mapsto \{0, 1\}$ checks whether f is satisfied in g
    - G_f enumerates all assignments of f,
    - $\sum_{g\in G_f} \phi_f(\mathbf{v}_g)$ measures the number of satisfied groundings of f.

However, MLN inference has been a challenging problem since 2006.

- Lifted inference falls short in handling distinctive evidence [5, 17, 13, 6, 8].
- In general, the direct inference is #P-complete [4].
- The most relevant works, pLogicNet and ExpressGNN [15, 23], used variational EM but the inference remains inefficient.

Our Approach: LogicMP

Motivation: More Efficient

- We use mean-field variational inference [24, 18, 11] to expand the MLN inference into forward computation.
- We use the structural symmetries in first-order logic for parallel computation.

LogicMP

Here, we present LogicMP, a method to encode first-order logic constraints over the neural network.

- It is valid for first-order logic.
- It is efficient using parallel computation.
- It is valid for arbitrary neural networks.

Approach Details - 1

- Recap the joint distribution with the neural network and the Markov logic network (MLN):
  - $p(\mathbf{v}|O) \propto \exp(\underbrace{\sum_i \phi_u(v_i)}_{Neural semantics} + \underbrace{\sum_{f \in F} w_f \sum_{g \in G_f} \phi_f(\mathbf{v}_g))}_{First-order logic}$

    where $\mathbf{v}$ is the set of unobserved variables. The second term is for symbolic FOLCs, where $\sum_{g \in G_f} \phi_f(\mathbf{v}_g))$ measures the number of satisfied groundings of f.

Approach Details - 2

- Perform mean-field variational inference over MLN.
  - $Q_i(v_i) \leftarrow \frac{1}{Z_i}\exp(\phi_u(v_i) + \sum_{f\in F} w_f \sum_{g\in G_f(i)} \hat{Q}_{i,g}(v_i))$ where Z_i is the partition function, $G_f(i)$ is the groundings of f that involve the ground atom i, and
  - $\hat{Q}_{i,g}(v_i) \leftarrow \sum_{\mathbf{v}_{g_{-i}}} \phi_f(v_i, \mathbf{v}_{g_{-i}}) \prod_{j\in g_{-i}} Q_j(v_j)$ is the grounding message that conveys information from the variables g_{-i} to the variable i w.r.t. the grounding g. g_{-i} denotes the ground atoms in g except i, e.g., $g_{-\mathtt{C}(e_1,e_3)} = \{\mathtt{C}(e_1,e_2), \mathtt{C}(e_2,e_3)\}$.

Approach Details - 3

- Less Computation per Grounding Message.
  - $Q_i(v_i) \leftarrow \frac{1}{Z_i} \exp(\phi_u(v_i) + \sum_{f \in F} w_f \sum_{g \in G_f(i)} \hat{Q}_{i,g}(v_i))$
  - $\cancel{\hat{Q}_{i,g}(v_i) \leftarrow \sum_{\mathbf{v}_{g_{-i}}} \phi_f(v_i, \mathbf{v}_{g_{-i}}) \prod_{j \in g_{-i}} Q_j(v_j)}$.
  - $\hat{Q}_{i,g}(v_i) \leftarrow 1_{v_i = \neg n_i} \prod_{j \in g_{-i}} Q_j(v_j = n_j)$ [Theorem 3.1]

Approach Details - 4

- Convert the inference into tensor parallel computations.
  - $\check{\mathbf{Q}}_{r_h}^{[f,h]}(\mathbf{v}_{r_h}) \leftarrow 1_{\mathbf{v}_{r_h}=\neg n_h}\texttt{einsum}(``\ldots, \mathcal{A}_{r_{j\neq h}}^{f}, \ldots \rightarrow \mathcal{A}_{r_h}^{f}", \ldots, \mathbf{Q}_{r_{j\neq h}}(n_{j\neq h}), \ldots)$
  - $\mathbf{Q}_r(\mathbf{v}_r) \leftarrow \frac{1}{\mathbf{Z}_r}\exp(\Phi_u(\mathbf{v}_r) + \sum_{[f,h],r=r_h} w_f \check{\mathbf{Q}}_{r_h}^{[f,h]}(\mathbf{v}_{r_h}))$

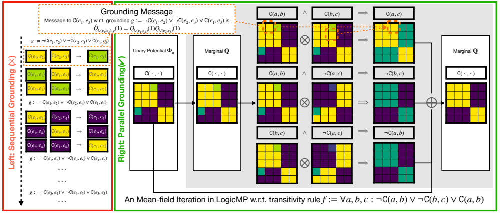

图 2: Instead of sequentially generating groundings (**left**), we exploit the structure of rules and formalize the MF iteration into Einstein summation notation, which enables parallel computation (**right**).

Experiments

Encoding FOLC over Document Images

- Task: Given the input image and input tokens, the task is to develop a function to predict whether two tokens coexist in a block.
- Rule: If tokens i and j are in the same block and tokens j and k are also together, then tokens i and k should be in the same block.

表 1: Comparison of F1 on FUNSD. Better results are in bold. “full” denotes the full set while “long” only considers the blocks with more than 20 tokens. “-” means failure.

Methods	full	long
LayoutLM-BIOES [22]	80.1	33.7
LayoutLM-SpanNER [7]	74.0	22.0
LayoutLM-SPADE [10]	80.1	43.5
LayoutLM-Pair [20]	82.0	46.7
LayoutLM-Pair w/ SL [21]	-	-
LayoutLM-Pair w/ SPL [1]	-	-
LayoutLM-Pair w/ SLrelax	82.0	47.8
LayoutLM-Pair w/ LogicMP	**83.3**	**50.1**
LayoutLM-Pair w/ SLrelax+LogicMP	**83.4**	**50.3**

Encoding FOLCs over Relational Graphs

- Task: Given the relational facts, the task is to develop a function to predict whether a latent fact is true.
- Rule: Rules of family/school/academic relations.

表 2: AUC-PR on Kinship, UW-CSE, and Cora. The best results are in bold. "-" means failure.

	Method	Kinship						UW-CSE						Cora					
		S1	S2	S3	S4	S5	avg.	A.	G.	L.	S.	T.	avg.	S1	S2	S3	S4	S5	avg.
MLN	MCMC [16]	.53	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
MLN	BP/Lifted BP [17]	.53	.58	.55	.55	.56	.56	.01	.01	.01	.01	.01	.01	-	-	-	-	-	-
MLN	MC-SAT [14]	.54	.60	.55	.55	-	-	.03	.05	.06	.02	.02	.04	-	-	-	-	-	-
MLN	HL-MRF [2]	**1.0**	**1.0**	**1.0**	**1.0**	-	-	.06	.09	.02	.04	.03	.05	-	-	-	-	-	-
NN+	ExpressGNN	.56	.55	.49	.53	.55	.54	.01	.01	.01	.01	.01	.01	.37	.66	.21	.42	.55	.44
NN+	ExpressGNN w/ GS [23]	.97	.97	.99	.99	.99	.98	.09	.19	.14	.06	.09	.11	.62	.79	.46	.57	.75	.64
NN+	ExpressGNN w/ LogicMP	.99	.98	**1.0**	**1.0**	**1.0**	**.99**	**.26**	**.30**	**.42**	**.25**	**.28**	**.30**	**.80**	**.88**	**.72**	**.83**	**.89**	**.82**

Encoding FOLCs over Text

- Task: Given the text sequence, the task is to develop a function to predict the sequence labels.
- Rule: adjacent rules and list rule.

表 3: Comparison of F1 on CoNLL2003. Better results are in bold. adj (list) denotes the adjacent (list) rules. “-” means failure.

Methods	F1
BLSTM [9]	89.98
BLSTM (lex) [3]	90.77
BLSTM w/ CRF [12]	90.94
BLSTM w/ CRF (mean field) [19]	91.07
BLSTM w/ SL [21]	-
BLSTM w/ SPL [1]	-
BLSTM w/ SLrelax	90.38
BLSTM w/ LogicDist (adj) [9]	p: 89.80, q: 91.11
BLSTM w/ LogicDist (adj+list) [9]	p: 89.93, q: 91.18
BLSTM w/ LogicMP (adj)	91.25
BLSTM w/ LogicMP (adj+list)	**91.42**

Conclusion

Conclusion

- LogicMP is an efficient MLN inference method.
- LogicMP is a neural layer with dense computations.
- LogicMP integrates FOLCs into any encoding network.
- LogicMP enjoys both the efficiency and effectiveness.

References i

K. Ahmed, S. Teso, K. Chang, G. V. den Broeck, and A. Vergari.
Semantic probabilistic layers for neuro-symbolic learning.
In Advances in Neural Information Processing Systems 35: Annual Conference on Neural Information Processing Systems 2022, NeurIPS 2022, New Orleans, LA, USA, November 28 - December 9, 2022, 2022.

S. H. Bach, M. Broecheler, B. Huang, and L. Getoor.
Hinge-loss Markov random fields and probabilistic soft logic.
The Journal of Machine Learning Research, 18:109:1–109:67, 2017.

References ii

- J. P. C. Chiu and E. Nichols.
  Named entity recognition with bidirectional LSTM-CNNs.
  Transactions of the Association for Computational Linguistics, 4:357–370, 2016.
- N. Dalvi and D. Suciu.
  The dichotomy of probabilistic inference for unions of conjunctive queries.
  Journal of the ACM (JACM), 59(6):1–87, 2013.

References iii

R. de Salvo Braz, E. Amir, and D. Roth.
Lifted first-order probabilistic inference.
In Proceedings of the Nineteenth International Joint Conference on Artificial Intelligence, Edinburgh, Scotland, UK, July 30 - August 5, 2005, pages 1319–1325. Professional Book Center, 2005.

G. V. den Broeck and J. Davis.
Conditioning in first-order knowledge compilation and lifted probabilistic inference.

In Proceedings of the Twenty-Sixth AAAI Conference on Artificial Intelligence, July 22-26, 2012, Toronto, Ontario, Canada. AAAI Press, 2012.

J. Fu, X. Huang, and P. Liu.
SpanNER: Named entity re-/recognition as span prediction.
In Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing, ACL/IJCNLP 2021, (Volume 1: Long Papers), Virtual Event, August 1-6, 2021, pages 7183–7195. Association for Computational Linguistics, 2021.

References v

E. Gribkoff, G. V. den Broeck, and D. Suciu.
Understanding the complexity of lifted inference and asymmetric weighted model counting.
In Statistical Relational Artificial Intelligence, Papers from the 2014 AAAI Workshop, Québec City, Québec, Canada, July 27, 2014, volume WS-14-13 of AAAI Technical Report. AAAI, 2014.

Z. Hu, X. Ma, Z. Liu, E. H. Hovy, and E. P. Xing.
Harnessing deep neural networks with logic rules.
In Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics, ACL 2016, August 7-12, 2016, Berlin, Germany, Volume 1: Long Papers. Association for Computer Linguistics, 2016.

References vii

W. Hwang, J. Yim, S. Park, S. Yang, and M. Seo.
Spatial dependency parsing for semi-structured document information extraction.
In Findings of the Association for Computational Linguistics: ACL/IJCNLP 2021, Online Event, August 1-6, 2021, volume ACL/IJCNLP 2021 of Findings of ACL, pages 330–343. Association for Computational Linguistics, 2021.

D. Koller and N. Friedman.
Probabilistic Graphical Models - Principles and Techniques.
MIT Press, 2009.

G. Lample, M. Ballesteros, S. Subramanian, K. Kawakami, and C. Dyer.
Neural architectures for named entity recognition.
In NAACL HLT 2016, The 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, San Diego California, USA, June 12-17, 2016, pages 260–270. Association for Computational Linguistics, 2016.

M. Niepert.
Lifted probabilistic inference: An MCMC perspective.
In 2nd International Workshop on Statistical Relational AI (StaRAI-12), held at the Uncertainty in Artificial Intelligence Conference (UAI 2012), Catalina Island, CA, USA, August 18, 2012, 2012.

H. Poon and P. M. Domingos.
Sound and efficient inference with probabilistic and deterministic dependencies.
In Proceedings, The Twenty-First National Conference on Artificial Intelligence and the Eighteenth Innovative Applications of Artificial

Intelligence Conference, July 16-20, 2006, Boston, Massachusetts, USA, pages 458–463. AAAI Press, 2006.

M. Qu and J. Tang.
Probabilistic logic neural networks for reasoning.
In Advances in Neural Information Processing Systems 32: Annual Conference on Neural Information Processing Systems 2019, NeurIPS 2019, December 8-14, 2019, Vancouver, BC, Canada, pages 7710–7720, 2019.

M. Richardson and P. M. Domingos.
Markov logic networks.
Machine Learning, 62(1-2):107–136, 2006.

P. Singla and P. M. Domingos.
Lifted first-order belief propagation.
In The Twenty-Third AAAI Conference on Artificial Intelligence, Chicago, Illinois, USA, July 13-17, 2008, pages 1094–1099. AAAI Press, 2008.

M. J. Wainwright and M. I. Jordan.
Graphical models, exponential families, and variational inference.
Foundations and Trends in Machine Learning, 1(1-2):1–305, 2008.

X. Wang, Y. Jiang, N. Bach, T. Wang, Z. Huang, F. Huang, and K. Tu.
AIN: fast and accurate sequence labeling with approximate inference network.
In Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing, EMNLP 2020, Online, November 16-20, 2020, pages 6019–6026. Association for Computational Linguistics, 2020.

J. Xu, W. Xu, M. Sun, T. Wang, and W. Chu.
Extracting trigger-sharing events via an event matrix.
In Findings of the Association for Computational Linguistics: EMNLP 2022, Abu Dhabi, United Arab Emirates, December 7-11, 2022, pages 1189–1201. Association for Computational Linguistics, 2022.

J. Xu, Z. Zhang, T. Friedman, Y. Liang, and G. V. den Broeck.
A semantic loss function for deep learning with symbolic knowledge.
In Proceedings of the 35th International Conference on Machine Learning, ICML 2018, Stockholmsmässan, Stockholm, Sweden, July

10-15, 2018, volume 80 of Proceedings of Machine Learning Research, pages 5498–5507. PMLR, 2018.

Y. Xu, M. Li, L. Cui, S. Huang, F. Wei, and M. Zhou.
Layoutlm: Pre-training of text and layout for document image understanding.
In KDD '20: The 26th ACM SIGKDD Conference on Knowledge Discovery and Data Mining, Virtual Event, CA, USA, August 23-27, 2020, pages 1192–1200. ACM, 2020.

References xv

Y. Zhang, X. Chen, Y. Yang, A. Ramamurthy, B. Li, Y. Qi, and L. Song.
Efficient probabilistic logic reasoning with graph neural networks.
In 8th International Conference on Learning Representations, Addis Ababa, Ethiopia, April 26-30, 2020. OpenReview.net, 2020.

S. Zheng, S. Jayasumana, B. Romera-Paredes, V. Vineet, Z. Su, D. Du, C. Huang, and P. H. S. Torr.
Conditional random fields as recurrent neural networks.
In 2015 IEEE International Conference on Computer Vision, Santiago, Chile, December 7-13, 2015, pages 1529–1537. IEEE Computer Society, 2015.