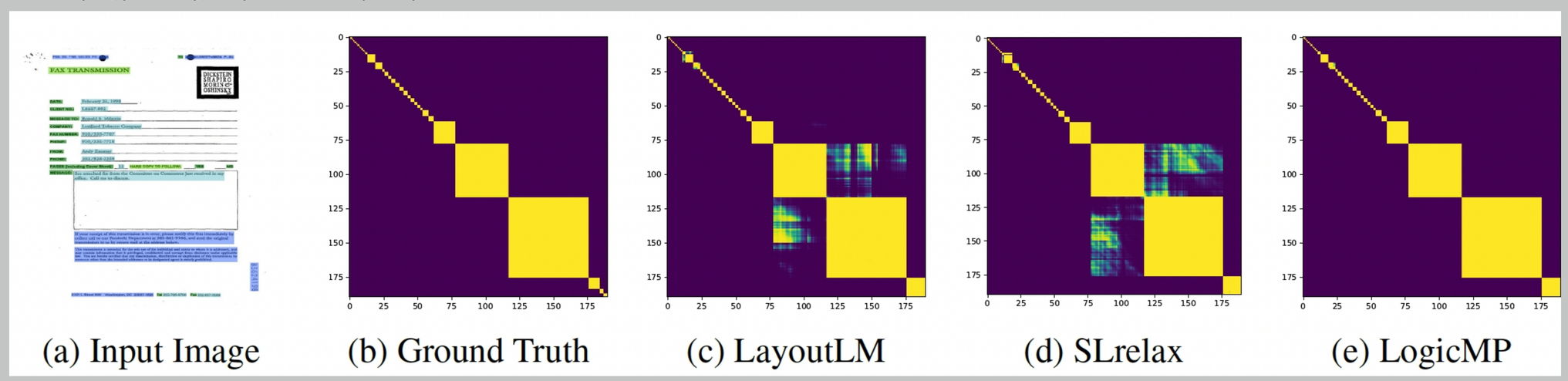
LogicMP: A Neuro-symbolic Approach for Encoding First-order Logic Constraints

Weidi Xu[1,2], Jingwei Wang[2], Lele Xie[2], Jianshan He[2], Hongting Zhou[2], Taifeng Wang[3], Xiaopei Wan[2], Jingdong Chen[2], Chao Qu[1], Wei Chu[2,1]

[1]INFLY TECH (Shanghai) Co., Ltd [2]Ant Group [3]BioMap Research

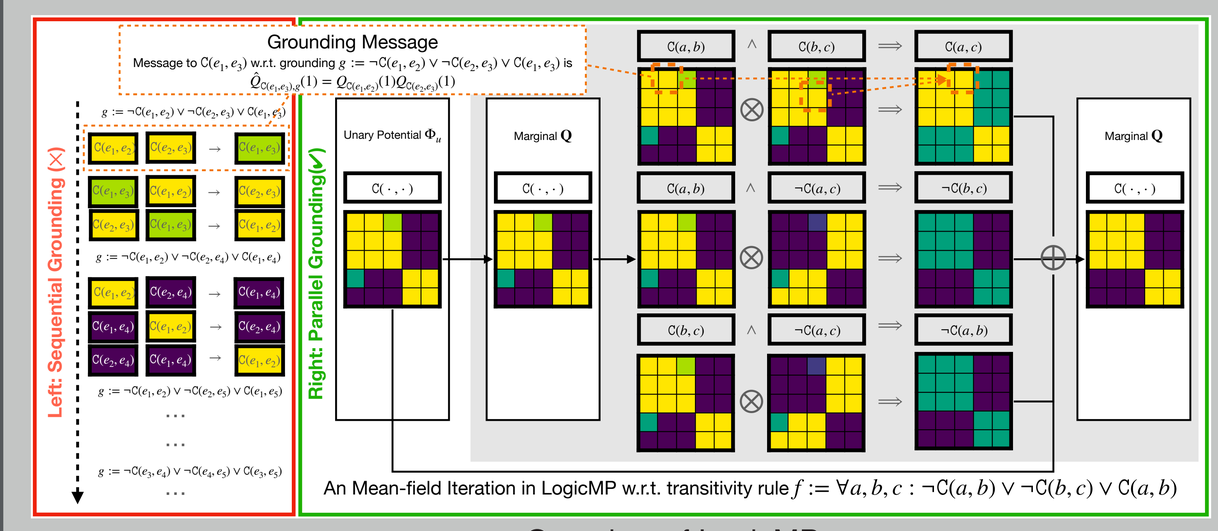
An Example of Encoding First-order Logic Constraints.

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.
Rule: $\forall \boldsymbol{i}, \boldsymbol{j}, \boldsymbol{k} : \mathtt{C}(\boldsymbol{i}, \boldsymbol{j}) \wedge \mathtt{C}(\boldsymbol{j}, \boldsymbol{k}) \implies \mathtt{C}(\boldsymbol{i}, \boldsymbol{k})$.



An example of using LogicMP in the image segmentation problem.

LogicMP: the first fully differentiable neuro-symbolic approach capable of encoding FOLCs for arbitrary neural networks



Overview of LogicMP.

Algorithm 2 PyTorch-like Code for LogicMP with Transitivity Rule

```
# logits:  torch.Tensor, size=[batchsize, nentities, nentities, 2]
# niterations:  int, number of iterations

cur_logits = logits.clone()
for i in range(niterations):
    Q = softmax(cur_logits, dim=-1)
    cur_logits = logits.clone()
    # Message Aggregation for Implication ∀a,b,c : C(a,b) ∧ C(b,c) ⟹ C(a,c)
    msg_to_ac = einsum('kab,kbc->kac', Q[...,1], Q[...,1])
    # Message Aggregation for Implication ∀a,b,c : C(a,b) ∧ ¬C(a,c) ⟹ ¬C(b,c)
    msg_to_bc = - einsum('kab,kac->kbc', Q[...,1], Q[...,0])
    # Message Aggregation for Implication ∀a,b,c : C(b,c) ∧ ¬C(a,c) ⟹ ¬C(a,b)
    msg_to_ab = - einsum('kbc,kac->kab', Q[...,1], Q[...,0])
    msg = msg_to_ac + msg_to_bc + msg_to_ac
    cur_logits[..., 1] += msg * weight
# Returns cur_logits
```

Pseudo Code for the Example.

Approach Details

Probabilistic modeling using Markov Logic Network (MLN)

$$p(\mathbf{v} \mid O) \propto \exp\Big(\underbrace{\sum_i \phi_u(v_i)}_{Neuralsemantics} + \underbrace{\sum_{f \in F} w_f \sum_{g \in G_f} \phi_f(\mathbf{v}_g)}_{First-orderlogic}\Big)$$

Performing mean-field variational inference over MLN.

$$Q_i(v_i) \leftarrow \frac{1}{Z_i} \exp\Big(\phi_u(v_i) + \sum_{f \in F} w_f \sum_{g \in G_f(i)} \hat{Q}_{i,g}(v_i)\Big)$$

$$\hat{Q}_{i,g}(v_i) \leftarrow \sum_{\mathbf{v}_{g_{-i}}} \phi_f(v_i, \mathbf{v}_{g_{-i}}) \prod_{j \in g_{-i}} Q_j(v_j)$$

Less computation per grounding message.

$$Q_i(v_i) \leftarrow \frac{1}{Z_i} \exp\Big(\phi_u(v_i) + \sum_{f \in F} w_f \sum_{g \in G_f(i)} \hat{Q}_{i,g}(v_i)\Big)$$

$$\hat{Q}_{i,g}(v_i) \leftarrow 1_{v_i = \neg n_i} \prod_{j \in g_{-i}} Q_j(v_j = n_j)$$

Converting the inference into tensor parallel computations.

$$\check{\mathbf{Q}}^{[f,h]}_{r_h}(\mathbf{v}_{r_h}) \leftarrow 1_{\mathbf{v}_{r_h} = \neg n_h} \mathtt{einsum}(\text{"}\ldots, \mathscr{A}^f_{r_{j \neq h}}, \ldots \rightarrow \mathscr{A}^f_{r_h}\text{"}, \ldots, \mathbf{Q}_{r_{j \neq h}}(n_{j \neq h}), \ldots)$$

$$\mathbf{Q}_r(\mathbf{v}_r) \leftarrow \frac{1}{\mathbf{Z}_r} \exp\Big(\Phi_u(\mathbf{v}_r) + \sum_{[f,h], r = r_h} w_f \check{\mathbf{Q}}^{[f,h]}_{r_h}(\mathbf{v}_{r_h})\Big)$$

Experiments

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.

Results over Document Images.

Methods	full	long
LayoutLM-BIOES [14]	80.1	33.7
LayoutLM-SpanNER [4]	74.0	22.0
LayoutLM-SPADE [6]	80.1	43.5
LayoutLM-Pair [12]	82.0	46.7
LayoutLM-Pair w/ SL [13]	-	-
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**

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.

Results over Relational Graphs.

	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 [9]	.53	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
	BP/Lifted BP [10]	.53	.58	.55	.55	.56	.56	.01	.01	.01	.01	.01	.01	-	-	-	-	-	-
	MC-SAT [8]	.54	.60	.55	.55	-	-	.03	.05	.06	.02	.02	.04	-	-	-	-	-	-
	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
	ExpressGNN w/ GS [15]	.97	.97	.99	.99	.99	.98	.09	.19	.14	.06	.09	.11	.62	.79	.46	.57	.75	.64
	ExpressGNN w/ LogicMP	.99	.98	**1.0**	**1.0**	**1.0**	**.99**	**.26**	**.30**	**.42**	**.25**	**.28**	**.30**	**.80**	**.88**	**.72**	**.83**	**.89**	**.82**

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

Results over Text.

Methods	F1
BLSTM [5]	89.98
BLSTM (lex) [3]	90.77
BLSTM w/ CRF [7]	90.94
BLSTM w/ CRF (mean field) [11]	91.07
BLSTM w/ SL [13]	-
BLSTM w/ SPL [1]	-
BLSTM w/ SLrelax	90.38
BLSTM w/ LogicDist (adj) [5]	p: 89.80, q: 91.11
BLSTM w/ LogicDist (adj+list) [5]	p: 89.93, q: 91.18
BLSTM w/ LogicMP (adj)	91.25
BLSTM w/ LogicMP (adj+list)	**91.42**

wdxu@inftech.ai

Weidi Xu[1,2], Jingwei Wang[2], Lele Xie[2], Jianshan He[2], Hongting Zhou[2], Taifeng Wang[3], Xiaopei Wan[2], Jingdong Chen[2], Chao Qu[1], Wei Chu[2,1]

[1]INFLY TECH (Shanghai) Co., Ltd [2]Ant Group [3]BioMap Research

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.
- 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.
- 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.
- 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.
- 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.
- 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.
- 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. 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.
- 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.
- 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.