Modeling Natural Language via Quantum Many-body Wave Function and Tensor Network

Tianjin University

Yitong Yao, Peng Zhang*, Jing Zhang
College of Intelligence and Computing, Tianjin University, China

Introduction

Quantum Many-body Wave Function inspired Language Modeling (QMWF-LM)

- regards different words as diverse particles in a many-body system, and a word containing multiple meanings equals to a particle lying in a superposition.
- Thus directly use the wave function to represent the words in order to express the complex interactions between words with different meanings.

Tensor Space Language Model (TSLM)

- TSLM is a generalization of the n-gram language model.
- a recursive calculation of conditional probability for language modeling is derived via tensor decomposition in TSLM.

QMWF-LM

- Product State representation for a sequence of words

$$|\psi_S^{ps}\rangle = |w_1\rangle \otimes \cdots \otimes |w_N\rangle$$

- Global representation:

$$|\psi_S\rangle = \sum_{h1\ldots hN}^{M} \mathcal{T}_{h_1\ldots h_N} |\phi_{h_1}\rangle \otimes \cdots \otimes |\phi_{h_N}\rangle$$

- Projection:

$$\langle\psi_S|\psi_S^{ps}\rangle = \sum_{h_1,\ldots,h_N}^{M} \mathcal{T}_{h_1\ldots h_N} \times \mathcal{A}_{h_1\ldots h_N} = \sum_{r=1}^{R} t_r \prod_{i=1}^{N} \left(\sum_{h_i=1}^{M} e_{r,i,h_i} \cdot \alpha_{i,h_i}\right)$$

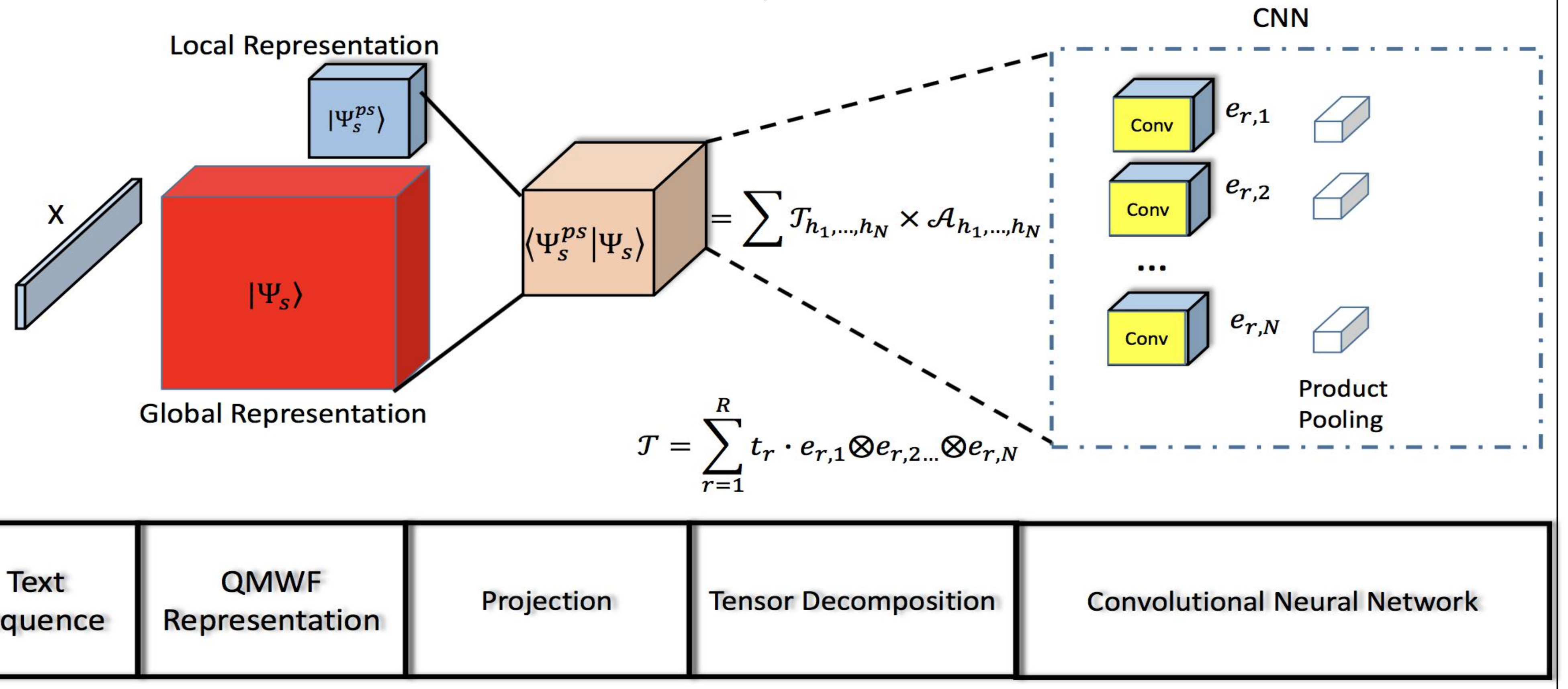


Table 1: Projection using CNN architecture

Input	$\boldsymbol{x}_i = (\alpha_{i,h_1}, \ldots, \alpha_{i,h_M})^T$
Convolution	$\Sigma_{r,i} = \sum_{h_i=1}^{M} e_{r,i,h_i} \cdot \alpha_{i,h_i}$
Product pooling	$\Pi_r = \prod_{i=1}^{N} \Sigma_{r,i}$
Output	$\sum_{r=1}^{R} t_r \cdot \Pi_r$

TSLM

- Sentence representation:

$$\boldsymbol{s} = \boldsymbol{w}_1 \otimes \cdots \otimes \boldsymbol{w}_n = \sum_{d_1,\cdots,d_n=1}^{m} \mathcal{A}_{d_1\cdots d_n} e_{d_1} \otimes \cdots \otimes e_{d_n}$$

- Corpus representation:

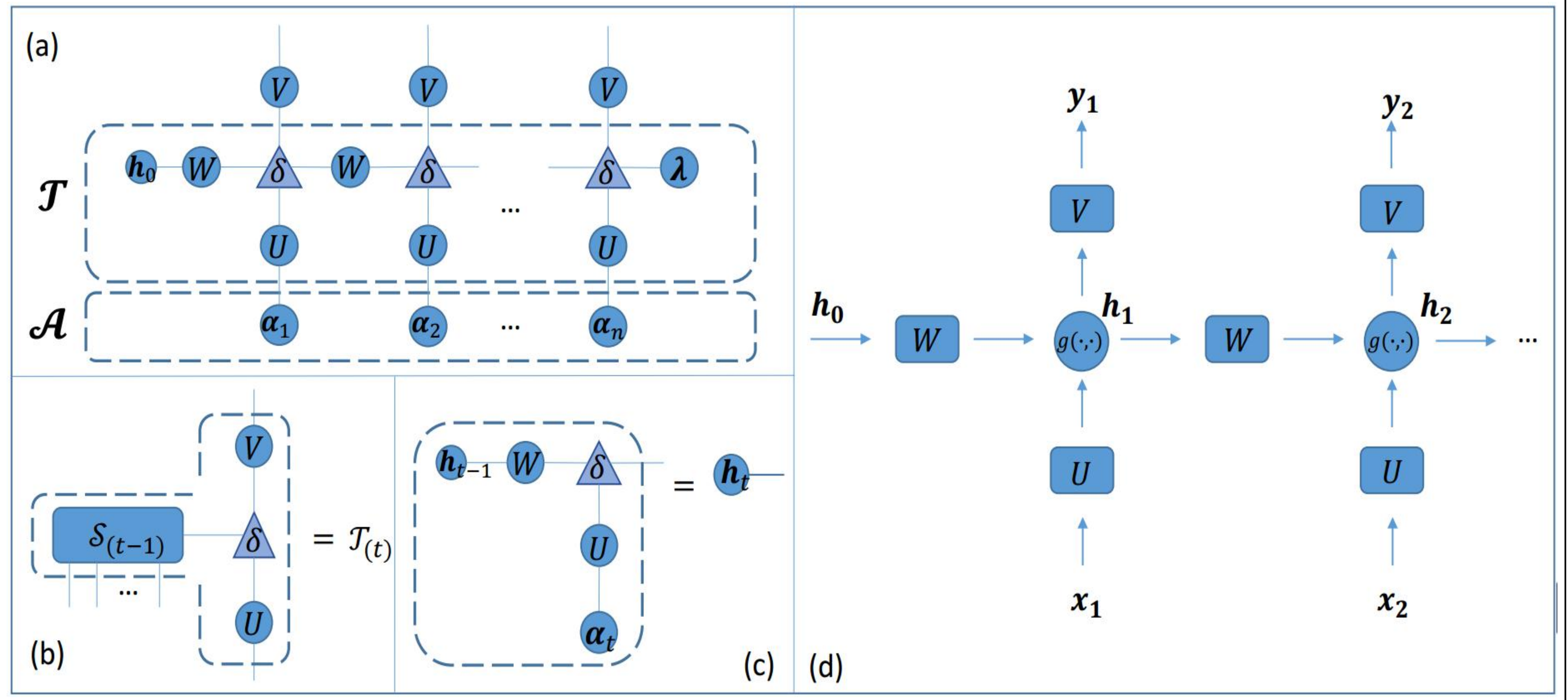
$$\boldsymbol{c} = \sum_i p_i s_i = \sum_{d_1,\cdots,d_n=1}^{m} \mathcal{T}_{d_1\ldots d_n} e_{d_1} \otimes \cdots \otimes e_{d_n}$$

- Sentence probability:

$$p(\boldsymbol{s}) = \langle \boldsymbol{s}, \boldsymbol{c} \rangle = \sum_{d_1,\cdots,d_n=1}^{m} \mathcal{T}_{d_1\ldots d_n} \mathcal{A}_{d_1\ldots d_n}$$

- The conditional probability $p(w_i|w_1^{i-1})$ can be computed as:

$$p(w_i|w_1^{i-1}) = \frac{p(w_1^i)}{p(w_1^{i-1})} = \frac{\langle \boldsymbol{w}_1^i, \boldsymbol{c} \rangle}{\langle \boldsymbol{w}_1^{i-1}, \boldsymbol{c} \rangle}$$



- Computing the conditional probability recursively:

$$p(w_t|w_1^{t-1}) = softmax(\boldsymbol{y}_t)$$
$$\boldsymbol{y}_t = V\boldsymbol{h}_t$$
$$\boldsymbol{h}_t = g(W\boldsymbol{h}_{t-1}, U\boldsymbol{\alpha}_t)$$
$$g(\boldsymbol{a}, \boldsymbol{b}) = \boldsymbol{a} \odot \boldsymbol{b}$$

Experimental Results

- Dataset: WIKIQA
- Model: QMWF-LM

Model	MAP	MRR
QLM	0.512	0.515
NNQLM-II	0.650	0.659
QA-CNN (Santos et al.) [10]	0.670	0.682
AP-CNN (Santos et al.) [10]	0.688	0.696
QMWF-LM-char	0.657 $^{\alpha}$	0.679 $^{\alpha}$
QMWF-LM-word	**0.695** $^{\alpha\beta}$	**0.710** $^{\alpha\beta}$

- Dataset: YahooQA
- Model: QMWF-LM

Model	P@1	MRR
Random guess	0.200	0.457
QLM	0.395	0.604
NNQLM-II	0.466	0.673
QA-CNN (Santos et al.)[10]	0.564	0.727
AP-CNN (Santos et al.)[10]	0.560	0.726
QMWF-LM-char	0.513 $^{\alpha\beta}$	0.696 $^{\alpha\beta}$
QMWF-LM-word	**0.575** $^{\alpha\beta}$	**0.745** $^{\alpha\beta}$

- Dataset: PTB, WikiText-2 Model: TSLM

	PTB				WikiText-2			
Model	Hidden size	Layers	Valid	Test	Hidden size	Layers	Valid	Test
KN-5(Mikolov and Zweig 2012)	-	-	-	141.2	-	-	-	-
RNN(Mikolov and Zweig 2012)	300	1	-	124.7	-	-	-	-
LSTM(Zaremba, Sutskever, and Vinyals 2014)	200	2	120.7	114.5	-	-	-	-
LSTM(Grave, Joulin, and Usunier 2016)	1024	1	-	82.3	1024	1	-	99.3
LSTM(Merity et al. 2017)	650	2	84.4	80.6	650	2	108.7	100.9
RNN†	256	1	130.3	124.1	512	1	126.0	120.4
LSTM†	256	1	118.6	110.3	512	1	105.6	101.4
TSLM	256	1	**117.2**	**108.1**	512	1	**104.9**	**100.4**
RNN+MoS†(Yang et al. 2018)	256	1	88.7	84.3	512	1	85.6	81.8
TSLM+MoS	256	1	**86.4**	**83.6**	512	1	**83.9**	**81.0**

Main References

[1] Zhang, Peng, et al. "A quantum many-body wave function inspired language modeling approach." *Proceedings of the 27th ACM International Conference on Information and Knowledge Management.* 2018.

[2] Zhang, Lipeng, et al. "A generalized language model in tensor space." *Proceedings of the AAAI Conference on Artificial Intelligence. Vol. 33. 2019.*

Conclusion

- QMWF-LM shows that it is feasible to use quantum theory to represent the interactions among words with multiple meanings.
- A generalized language model (TSLM) was introduced inspired by QMWF-LM, adopting the same method to express the word using quantum wave function, deriving a recursive representation of the conditional probability distribution