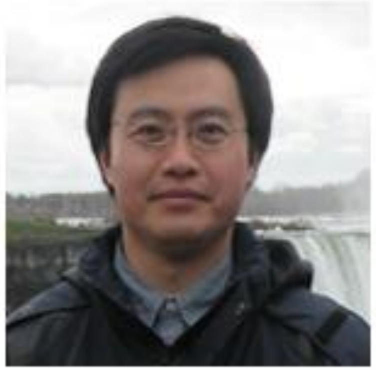
UNIVERSITÀ DEGLI STUDI DI PADOVA

Exploring **Interpretable** Neural Network by Quantum representation

Benyou Wang, Qiuchi Li, Prayag Tiwari, Massimo Melucci

University of Padova

Done with the collaboration



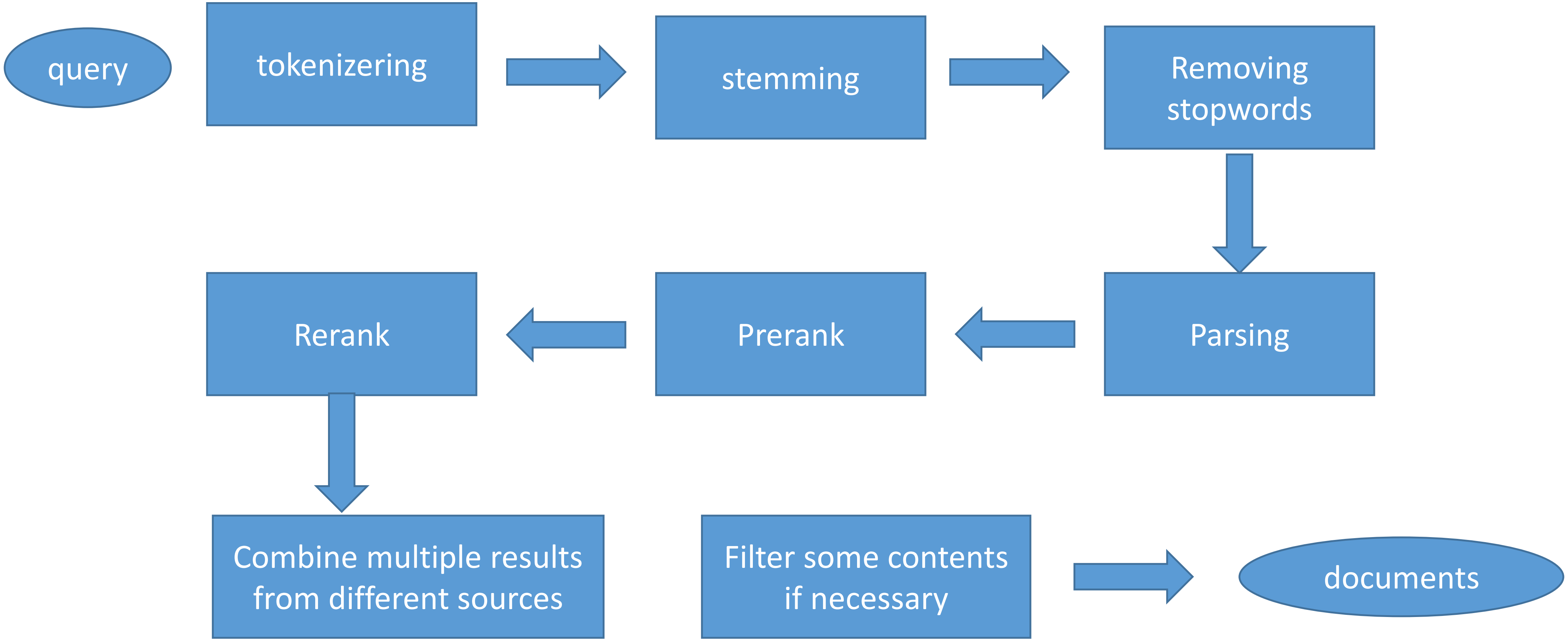
What is **Interpretability**

- Post-hoc explanations
  - Take a **learned model** and draw some kind of useful insights
  - E.g. Visualization in machine translation [Liu Yang &Maosong Sun ACL 2017]

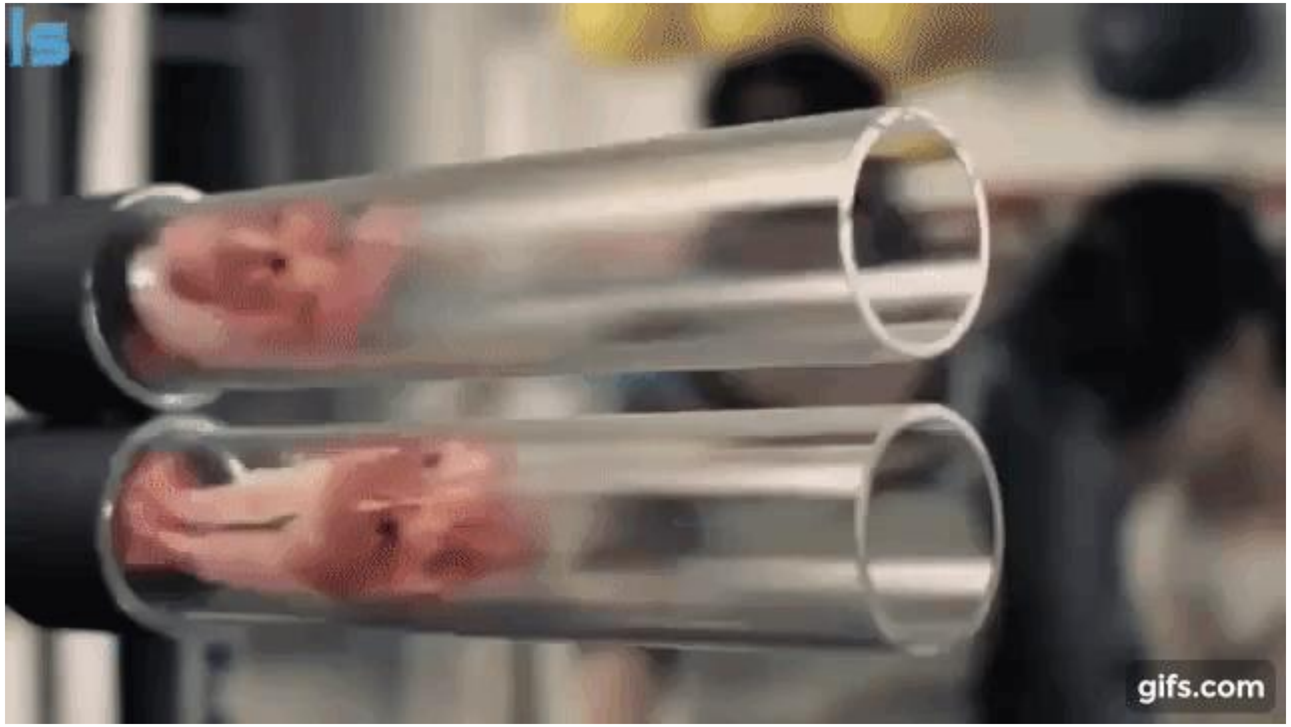
- Transparency
  - Targeting ``how does the model work?'' and seeks to provide some way to understand the core mechanisms
  - E.g. Capsule Network [Hinton NIPS 2017]

Zachary C Lipton. The mythos of model interpretability. arXiv preprint arXiv:1606.03490, 2016, **ICML** Workshop on Human Interpretability in Machine Learning
Yanzhuo Ding, Yang Liu, Huanbo Luan, and Maosong Sun. Visualizing and understanding neural machine translation. **ACL**, volume 1, pages 1150–1159, 2017.
Sabour S, Frosst N, Hinton G E. Dynamic routing between capsules[C]//**NIPS . 2017**: 3856-3866.

An **Pipeline** example for text processing


query
tokenizering
stemming
Removing stopwords
Parsing
Prerank
Rerank
Combine multiple results from different sources
Filter some contents if necessary
documents

Transparency in end-to-end Paradigm



https://www.youtube.com/watch?v=TYpBJ71VW9g

End to end mechanism

- ✓ Less accumulating error
- ✓ Less involvement with Human beings
- ✓ Improve performance with shared features of the downstream tasks and upstream tasks

- ❖ Hard to adjust
- ❖ Hard to transfer
- ❖ Hard to understand

We need End to End mechanism, but in a fine-grained way

Design each subcomponents in the End-2-end architecture with a good background of the task

- *Both language understanding and artificial intelligence require being able to understand bigger things from knowing about smaller parts*

Christopher Manning 2017

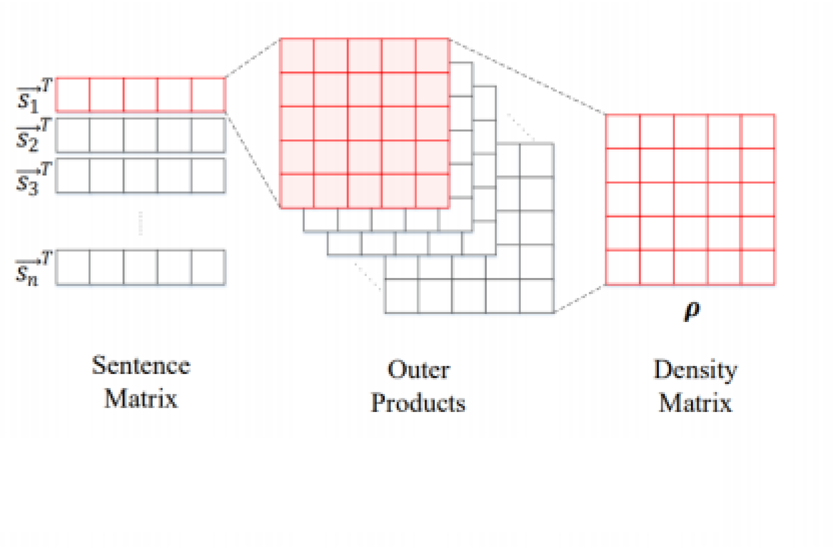
Motivations

- Design self-***explainable*** subcomponents in end2end network
- Provides more **transparency** from the network
- **Theoretical** explanations for why neural network works or why it dose not work

Contents

- End to End language model for QA [AAAI 2018]
- Quantum Many body function for language model in QA **[CIKM 2018]**
- Quantum-inspired word Embedding [ACL REP4NLP 2018]
- **Hibert Semantic Space [In process]**

End-2-end Language model for QA

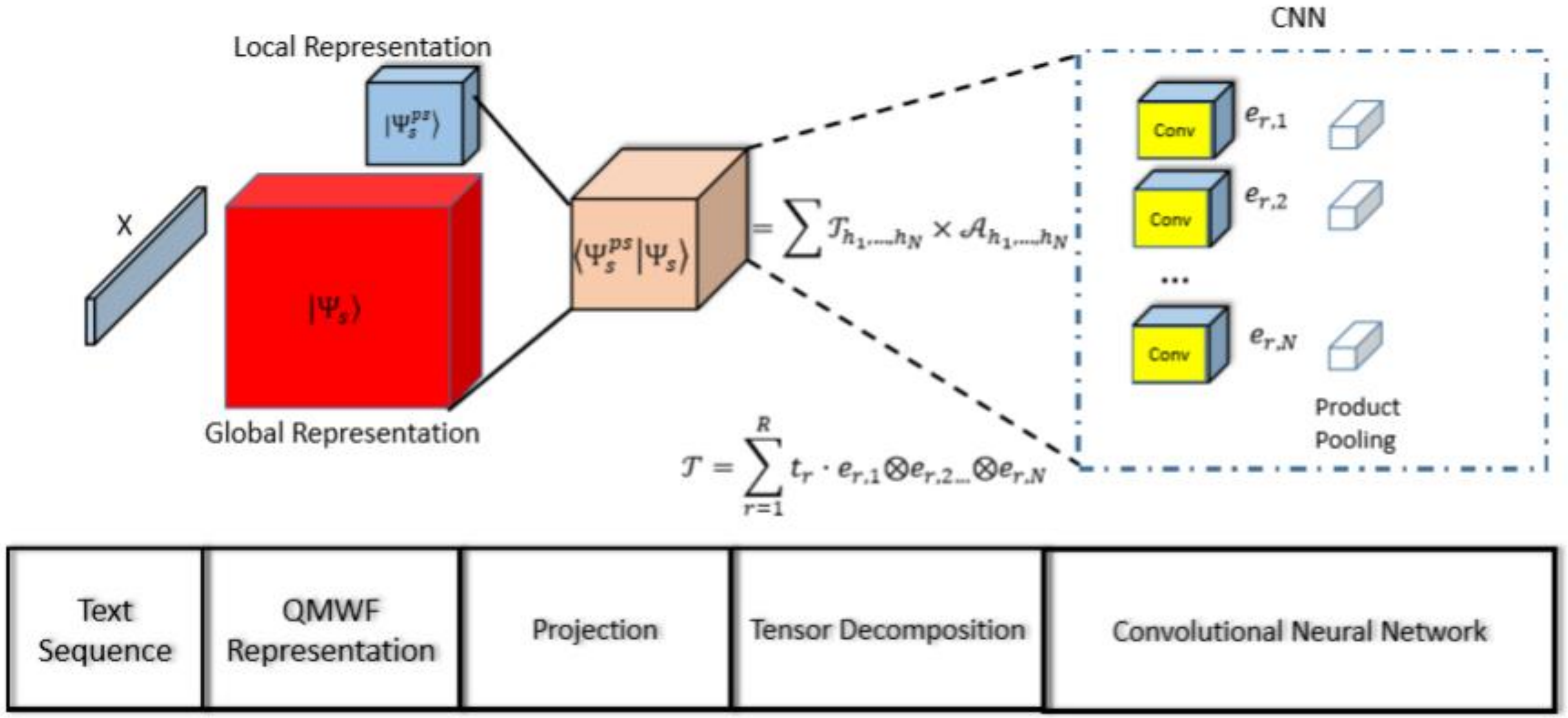


Matching with two matrices

- $tr(\rho_1\rho_2)$
- CNN over $\rho_1\rho_2$

Zhang Peng, Niu Jiabing, Su Zhan, **Wang Benyou** et al. End-to-End Quantum-like Language Models with Application to Question Answering **AAAI 2018**

Quantum many-body function for LM


CNN
Local Representation
$|\Psi_s^{ps}\rangle$
X
$|\Psi_s\rangle$
Global Representation
$\langle\Psi_s^{ps}|\Psi_s\rangle$
$= \sum \mathcal{T}_{h_1,\dots,h_N} \times \mathcal{A}_{h_1,\dots,h_N}$
Conv
$e_{r,1}$
Conv
$e_{r,2}$
...
Conv
$e_{r,N}$
Product Pooling
$\mathcal{T} = \sum_{r=1}^{R} t_r \cdot e_{r,1} \otimes e_{r,2\dots} \otimes e_{r,N}$
Text Sequence
QMWF Representation
Projection
Tensor Decomposition
Convolutional Neural Network

Use CNN to ***approximate*** Tensor Decomposition in the projection of Quantum Many-Body Language Function

Peng Zhang, Zhan Su, Lipeng Zhang, **Benyou Wang** , Dawei Song. 2018. A Quantum Many-body Wave Function Inspired Language Modeling Approach, **CIKM 2018**

Complex word-embedding

- Super-linearity superposition with phase

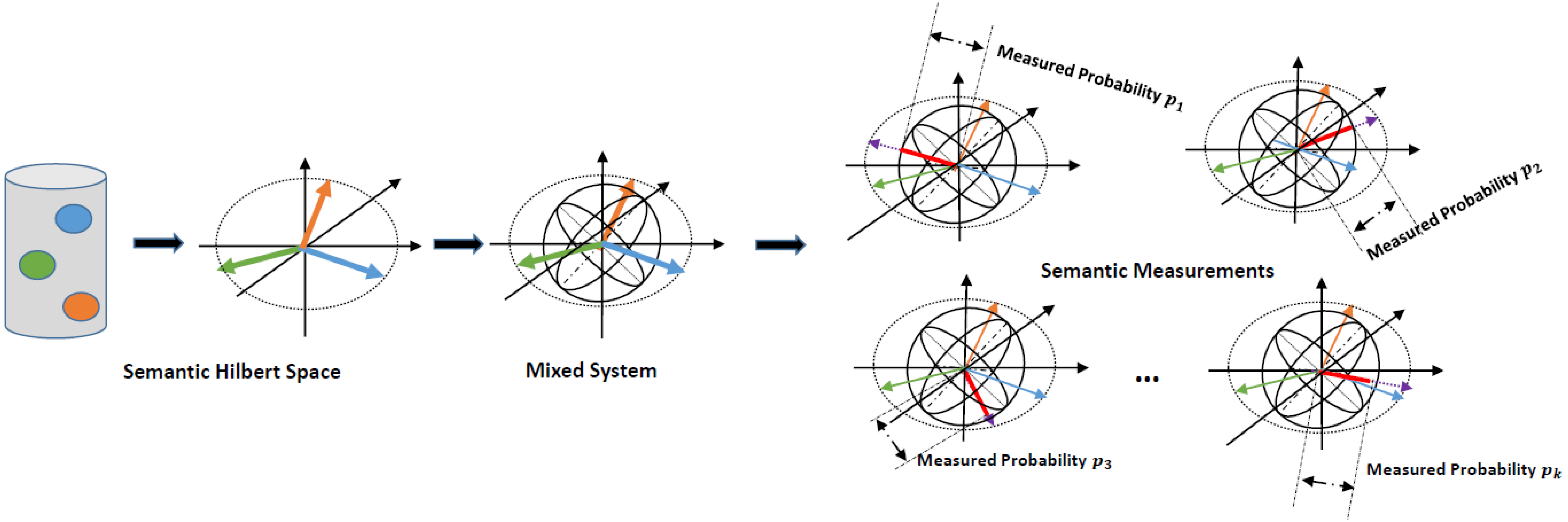
$$
\begin{aligned}
z^* &= z_1 + z_2 = r_1 e^{i\theta_1} + r_2 e^{i\theta_2} \\
&= \sqrt{r_1^2 + r_2^2 + 2r_1 r_2 \cos(\theta_2 - \theta_1)} \times e^{i \arctan\left(\frac{r_1 \sin(\theta_1) + r_2 \sin(\theta_2)}{r_1 \cos(\theta_1) + r_2 \cos(\theta_2)}\right)}
\end{aligned}
$$

Li Qiuchi, Uprety Sagar, **Wang Benyou** , Song Dawei Quantum-inspired Complex Word Embedding, ACL 2018 3rd Workshop on Representation Learning for NLP , **ACL 2018 RepL4NLP**

Hibert Semantic Space

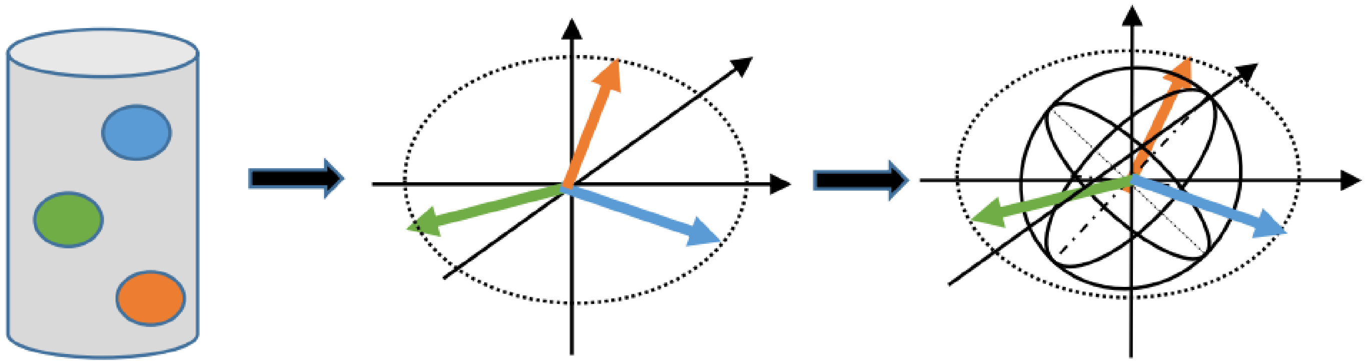
- Unify these four things in a complex-valued space
  - Semeses
  - Word
  - Phrase/Sentence/Documents
  - Topic as measurements

Framework


Semantic Hilbert Space
Mixed System
Measured Probability p_1
Measured Probability p_2
Semantic Measurements
Measured Probability p_3
...
Measured Probability p_k

Definition

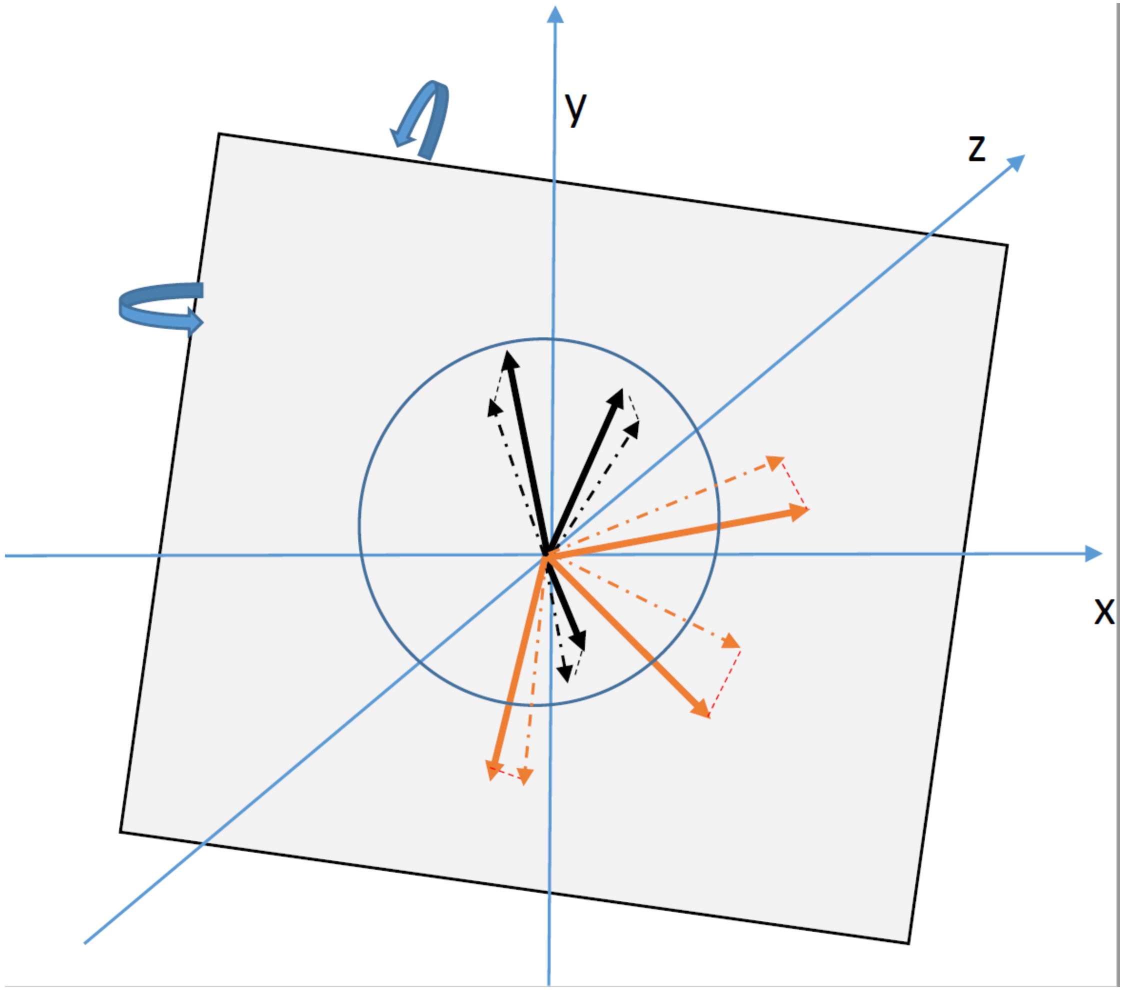
- Semeses as basic state
- Word as superstition state
- Sentence as mixed system



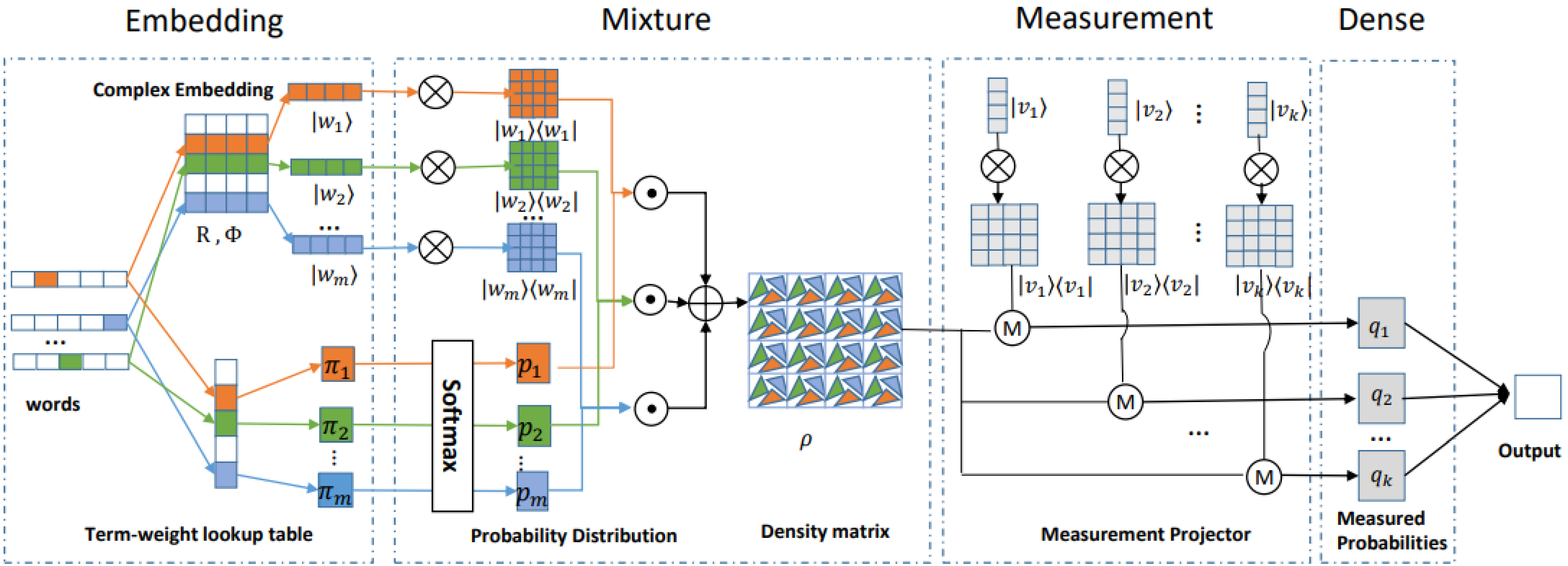
Semantic Hilbert Space

Mixed System

Trainable Measurements for sentence classification



Implements


Embedding
Mixture
Measurement
Dense
Complex Embedding
$|w_1\rangle$
$|w_2\rangle$
$|w_m\rangle$
R , Φ
words
π_1
π_2
π_m
Term-weight lookup table
$|w_1\rangle\langle w_1|$
$|w_2\rangle\langle w_2|$
$|w_m\rangle\langle w_m|$
Softmax
p_1
p_2
p_m
Probability Distribution
ρ
Density matrix
$|v_1\rangle$
$|v_2\rangle$
$|v_k\rangle$
$|v_1\rangle\langle v_1|$
$|v_2\rangle\langle v_2|$
$|v_k\rangle\langle v_k|$
M
Measurement Projector
q_1
q_2
q_k
Measured Probabilities
Output

Physical meaning for our models

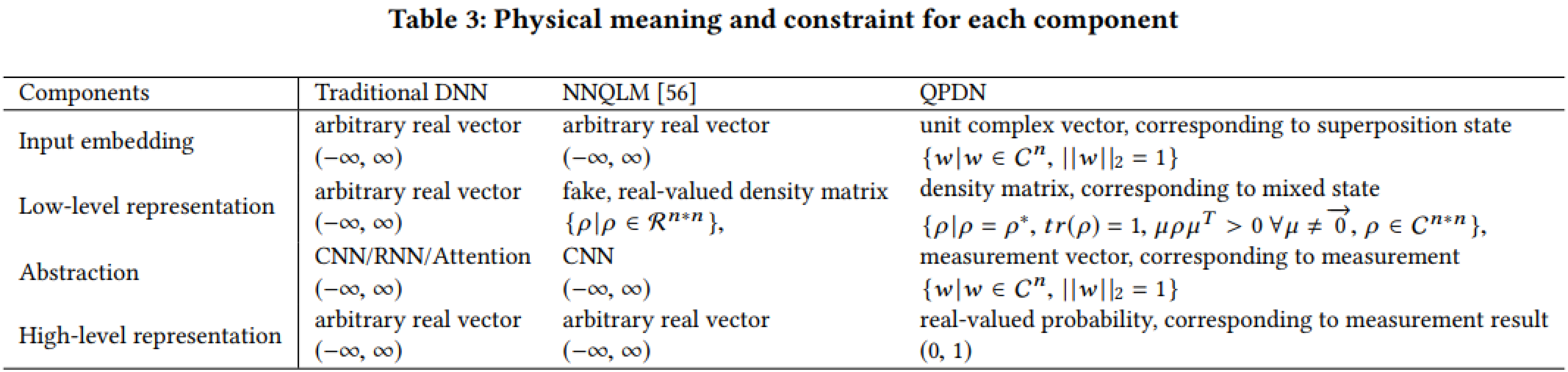


Table 3: Physical meaning and constraint for each component

Components	Traditional DNN	NNQLM [56]	QPDN
Input embedding	arbitrary real vector $(-\infty, \infty)$	arbitrary real vector $(-\infty, \infty)$	unit complex vector, corresponding to superposition state $\{w\|w \in C^n, \|\|w\|\|_2 = 1\}$
Low-level representation	arbitrary real vector $(-\infty, \infty)$	fake, real-valued density matrix $\{\rho\|\rho \in \mathcal{R}^{n*n}\}$,	density matrix, corresponding to mixed state $\{\rho\|\rho = \rho^*, tr(\rho) = 1, \mu\rho\mu^T > 0\ \forall\mu \neq \overrightarrow{0}, \rho \in C^{n*n}\}$,
Abstraction	CNN/RNN/Attention $(-\infty, \infty)$	CNN $(-\infty, \infty)$	measurement vector, corresponding to measurement $\{w\|w \in C^n, \|\|w\|\|_2 = 1\}$
High-level representation	arbitrary real vector $(-\infty, \infty)$	arbitrary real vector $(-\infty, \infty)$	real-valued probability, corresponding to measurement result $(0, 1)$

Experiments

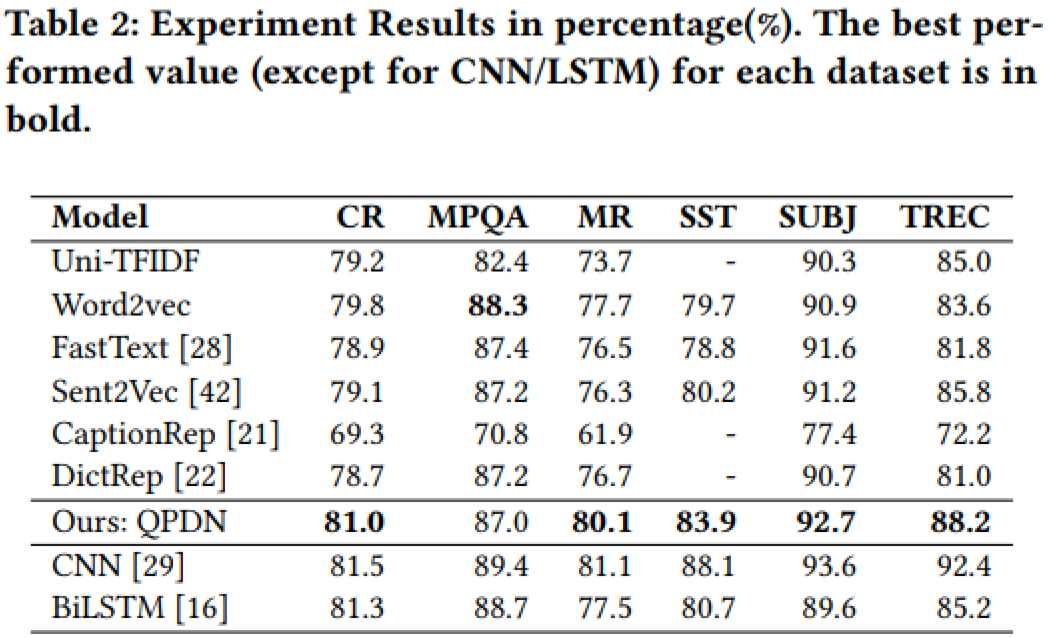


Table 2: Experiment Results in percentage(%). The best performed value (except for CNN/LSTM) for each dataset is in bold.

Model	CR	MPQA	MR	SST	SUBJ	TREC
Uni-TFIDF	79.2	82.4	73.7	-	90.3	85.0
Word2vec	79.8	**88.3**	77.7	79.7	90.9	83.6
FastText [28]	78.9	87.4	76.5	78.8	91.6	81.8
Sent2Vec [42]	79.1	87.2	76.3	80.2	91.2	85.8
CaptionRep [21]	69.3	70.8	61.9	-	77.4	72.2
DictRep [22]	78.7	87.2	76.7	-	90.7	81.0
Ours: QPDN	**81.0**	87.0	**80.1**	**83.9**	**92.7**	**88.2**
CNN [29]	81.5	89.4	81.1	88.1	93.6	92.4
BiLSTM [16]	81.3	88.7	77.5	80.7	89.6	85.2

Case study for our measurement

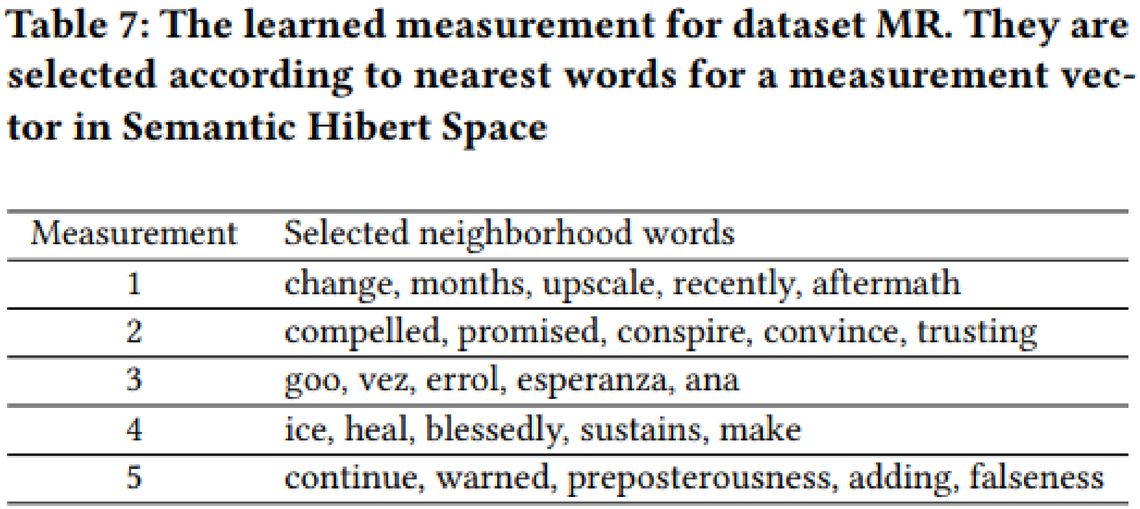


Table 7: The learned measurement for dataset MR. They are selected according to nearest words for a measurement vector in Semantic Hibert Space

Measurement	Selected neighborhood words
1	change, months, upscale, recently, aftermath
2	compelled, promised, conspire, convince, trusting
3	goo, vez, errol, esperanza, ana
4	ice, heal, blessedly, sustains, make
5	continue, warned, preposterousness, adding, falseness

Conclusion

- More concrete physical meaning
- Self-explainable subcomponents
- More constrain for the subcomponents
- Guided by Quantum probability theory

Future works with this topic

- Explore high-dimension **tensor network** with Quantum representation

- Capsule Network with Quantum insights