DEPARTMENT OF INFORMATION ENGINEERING
UNIVERSITY OF PADOVA

QUARTZ
Quantum Information Access and Retrieval Theory

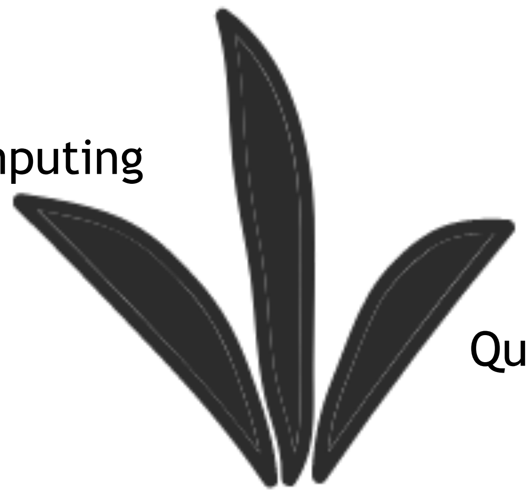
Quantum-Inspired IR/Language Modelling

Qiuchi Li, **Benyou Wang**

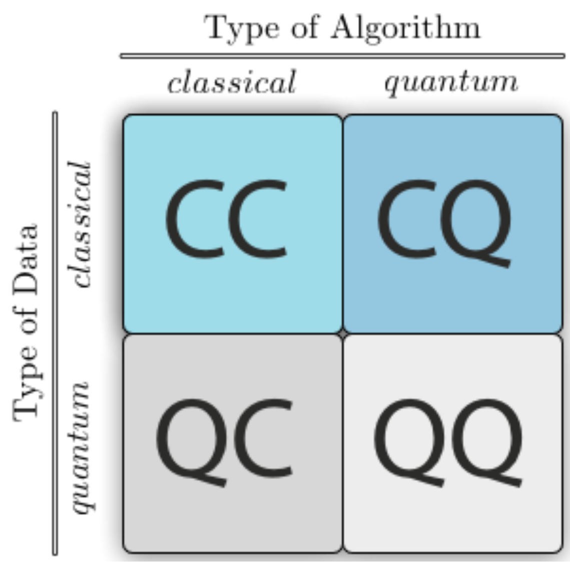
University of Padua

Toutiao, Beijing, China, 28/06/2019

Relationship with Quantum Computing



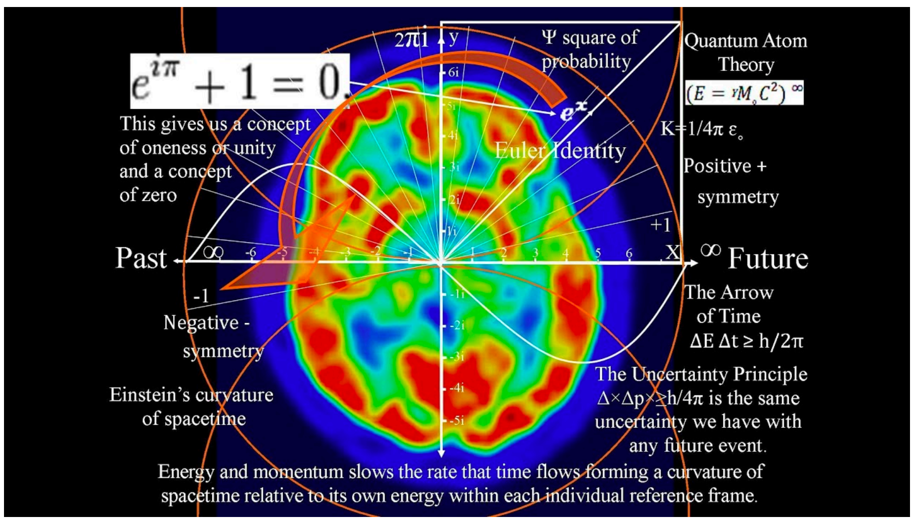
Relationship with **Quantum Machine Learning**



- Classical Machine learning from Quantum, e.g. Boltzmann machine, Gradient decent
- Machine Learning deployed on **Quantum computers, s**peed up the classical machine learning algorithms

Quantum theory Outside Physics

Quantum mind/brain/consciousness/cognition



- Social science
  - [E. Haven and A. Khrennikov. 2013. Quantum Social Science. Cambridge University Press.]
- Cognition science
  - [Jerome R. Busemeyer and Peter D. Bruza. 2013. Quantum Models of Cognition and Decision. Cambridge University Press]
- **Information retrieval**
  - [Alessandro Sordoni, Jian-Yun Nie, and Yoshua Bengio. 2013. Modeling term dependencies with quantum language models for IR. In Proc. of SIGIR. ACM, 653–662.]

- Our works do not rely on quantum cognition

Contents

- **History of quantum-inspired IR/NLP**
- Basics from Quantum Theory
- Semantic Hilbert Space—NAACL best paper
- Future work with Quantum Theory

Main Researchers

- Massimo Melucci, University of Padova
- Peng Zhang/Yuexian Hou,Tianjin University
- Dawei Song, BIT
- Christina Lioma, University of Copenhagen
- Peter Bruza, Queensland University of Technology
- Van Rijsbergen
- Diedarik Aerts and Andrei Khrennikov, Bruseymer.
- Jian-yun Nie, University of Montreal
- Ingo, lefist, guido zuccon, piwosiki

- SIGIR Shannon award
- ECIR
- ICTIR
- ICITR
- NAACL
- SIGIR
- ICTIR

Quantum IR

- *Quantum Formulism can formulate the different IR models (**logic, vector, probabilistic**, etc.) in a unified framework.*

C.J. van Rijsbergen
UK Royal Academy of Engineering Fellow
SIGIR 2006 Salton Award Lecture

[C.J. van Rijsbergen 2004, Geometry of Information Retreival]

[Piwowarski B, et al. What can quantum theory bring to information retrieval. CIKM 2010. 59–68]

Roadmap of Quantum IR formal models

Milestones

Quantum Analogy based IR Methods

Double Slit
(Zuccon et al. ECIR 2009)

Photon Polarization
(**Zhang**, et al. ECIR 2011, ICTIR 2011)

Pros & Cons:

+ Novel intuitions
[ECIR'11 Best Poster Award]

- Shallow analogy
- Inconsistent with quantum axioms

Quantum Language Models (QLMs)

Original QLM
(Sordoni et al. SIGIR 2013)

QLM variants
(Li, Li, **Zhang**, SIGIR 2015)
(Xie, Hou*, **Zhang***, IJCAI 2015)

Pros & Cons:

+ Consistent with axioms

- QLM components are designed separately, instead of learned jointly

Neural Quantum Language Models

End2end QLM for QA
(**Zhang** et al. AAAI 2018)

Further variants
(**Zhang** et al. Science China 2018)

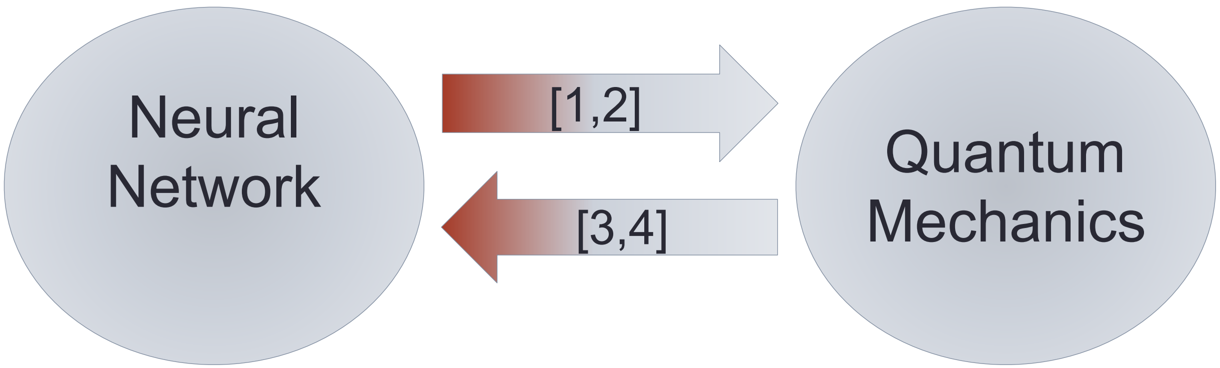
Pros & Cons:

+ Effective joint learning for Question answering;

- Lacks inherent connection between NN and QLM
- Cannot model complex interaction among words

Credits from Prof. Peng Zhang

Quantum Theory & NN



[1] Carleo G, Troyer M. Solving the quantum many-body problem with artificial neural networks[J]. **Science**, 2017, 355(6325): 602-606.
[2] Gao X, Duan L M. Efficient representation of quantum many-body states with deep neural networks[J]. **Nature communications**, 2017, 8(1): 662.
[3] Lin X, Rivenson Y, Yardimci N T, et al. All-optical machine learning using diffractive deep neural networks[J]. **Science**, 2018, 361(6406): 1004-1008.
[4] Levine Y, Yakira D, Cohen N, et al. Deep Learning and Quantum Entanglement: Fundamental Connections with Implications to Network Design[C]. **ICLR** 2018.

Contents

- History of quantum inspired IR/NLP
- **Basics from Quantum Theory**
- Semantic Hilbert Space—NAACL best paper
- Future work with Quantum Theory

Basic Dirac notations

- Bra & Ket
  - **Bra:** $< \cdot |$ like a **row vector**, e.g. $< x|$
  - **Ket**: $| \cdot >$ like a **column vector**, e.g., $|x >$

- Inner Product

$$< x|x >$$

- Outer Product

$$|x > < x|$$

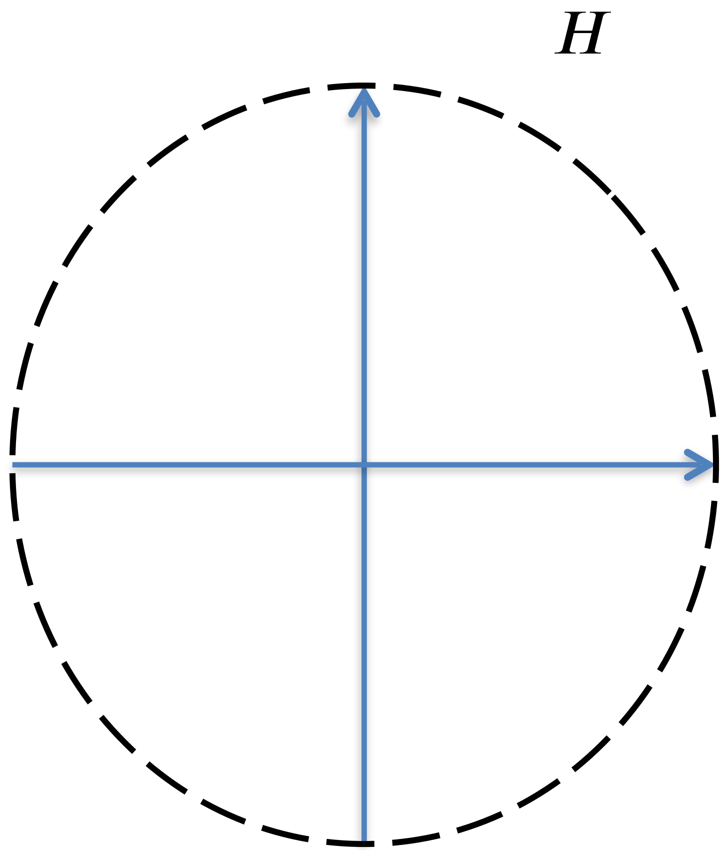
Four axioms in Quantum Mechanics

- **Axiom 1: State and Superposition**
- **Axiom 2: Measurements**
- Axiom 3: Composite system
- Axiom 4: Unitary Evolution

[Nielsen M A, Chuang I L. 2000]

Quantum Preliminaries

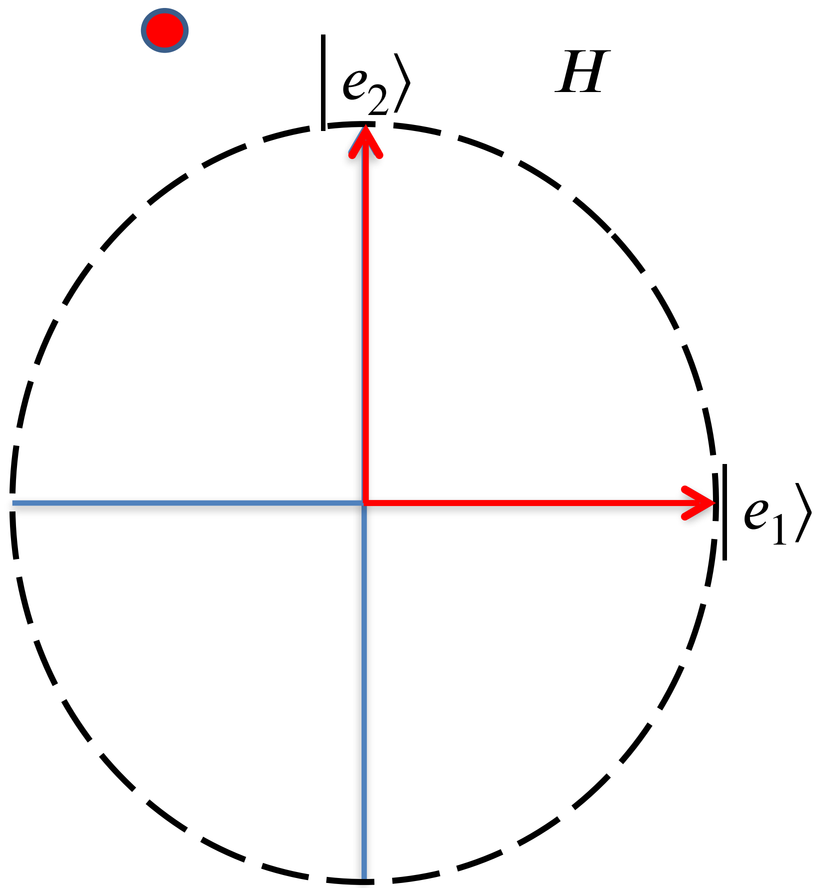
- **Hilbert Space**



Quantum Preliminaries

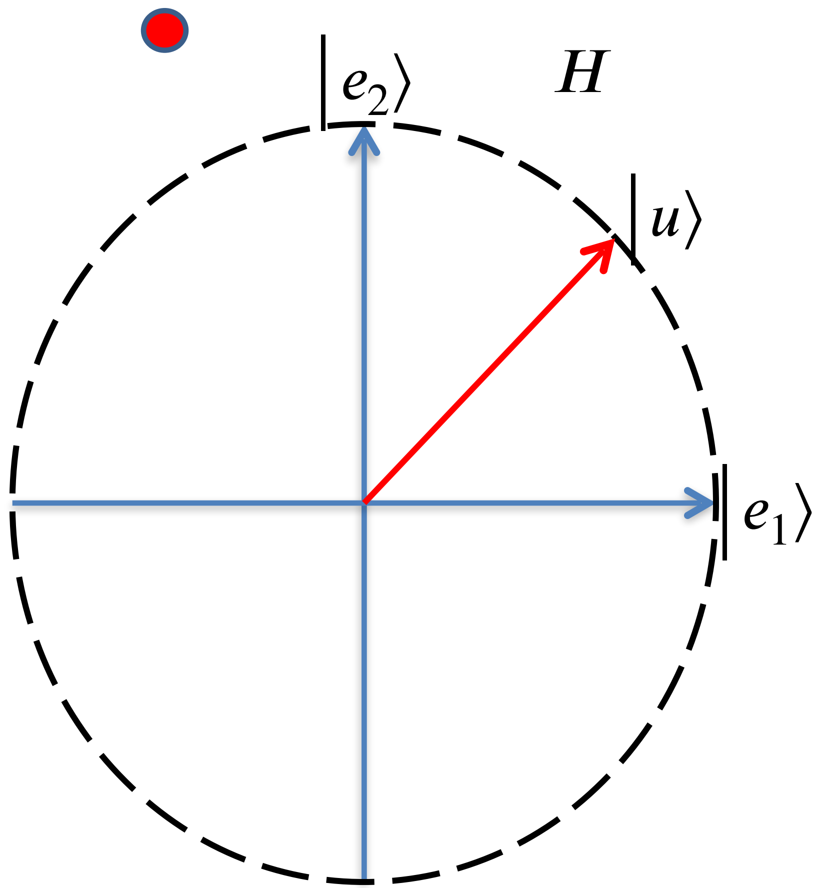
- Hilbert Space

- **Pure State**
  - **Basis State**

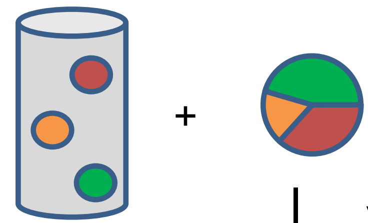


Quantum Preliminaries

- Hilbert Space
- **Pure State**
  - Basis State
  - **Superposition State**



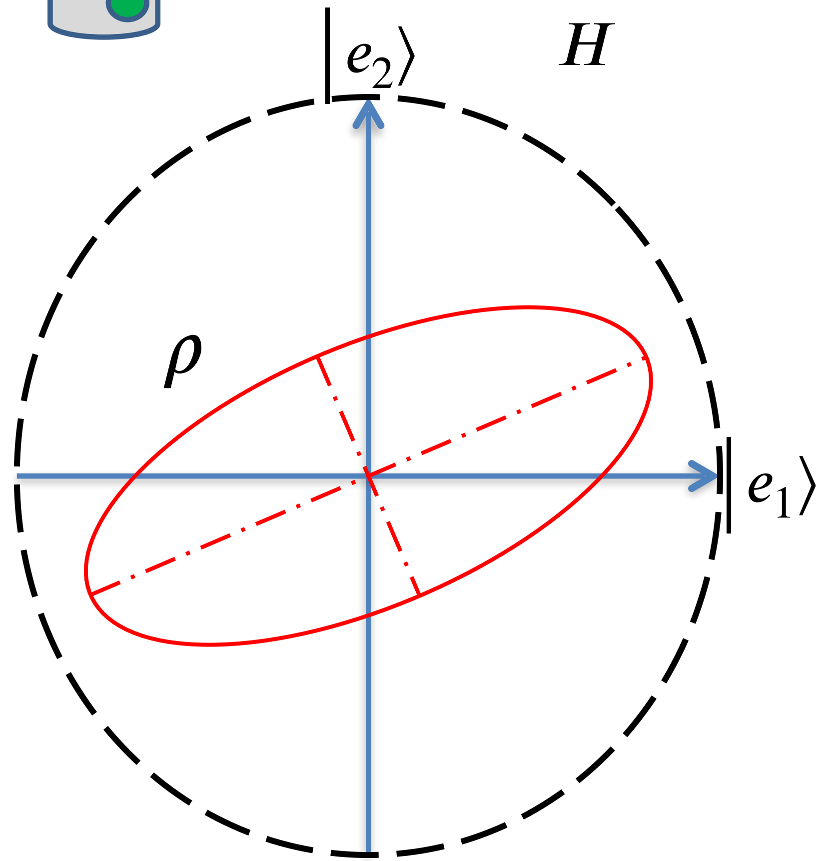
Quantum Preliminaries



- Hilbert Space

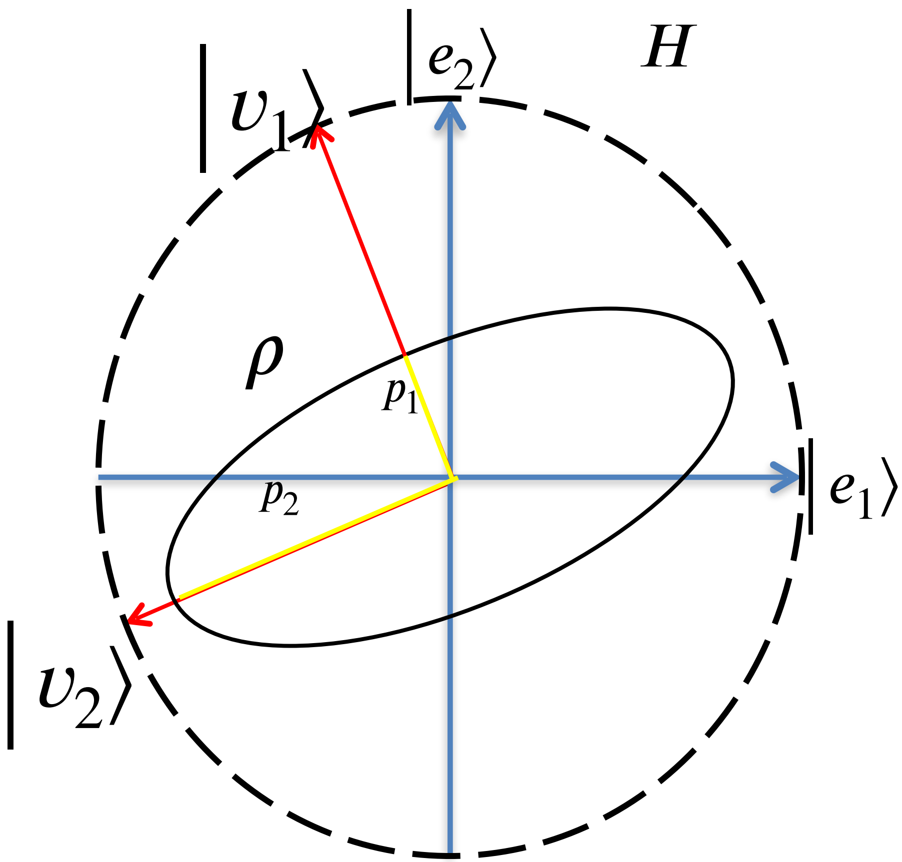
- Pure State
  - Basis State
  - Superposition State

- **Mixed State**



Quantum Preliminaries

- Hilbert Space
- Pure State
  - Basis State
  - Superposition State
- Mixed State
- **Measurement**



Contents

- History of quantum inspired IR/NLP
- Basics from Quantum theory
- **Semantic Hilbert Space—NAACL best paper**
- Future work with Quantum Theory

Research Problem

- Interpretability issue for NN-based NLP models

  1. **Transparency**: explainable component in the design phase

  2. **Post-hoc Explainability**: why the model works after execution

The Mythos of Model Interpretability, Zachery C. Lipton, 2016

Research questions :

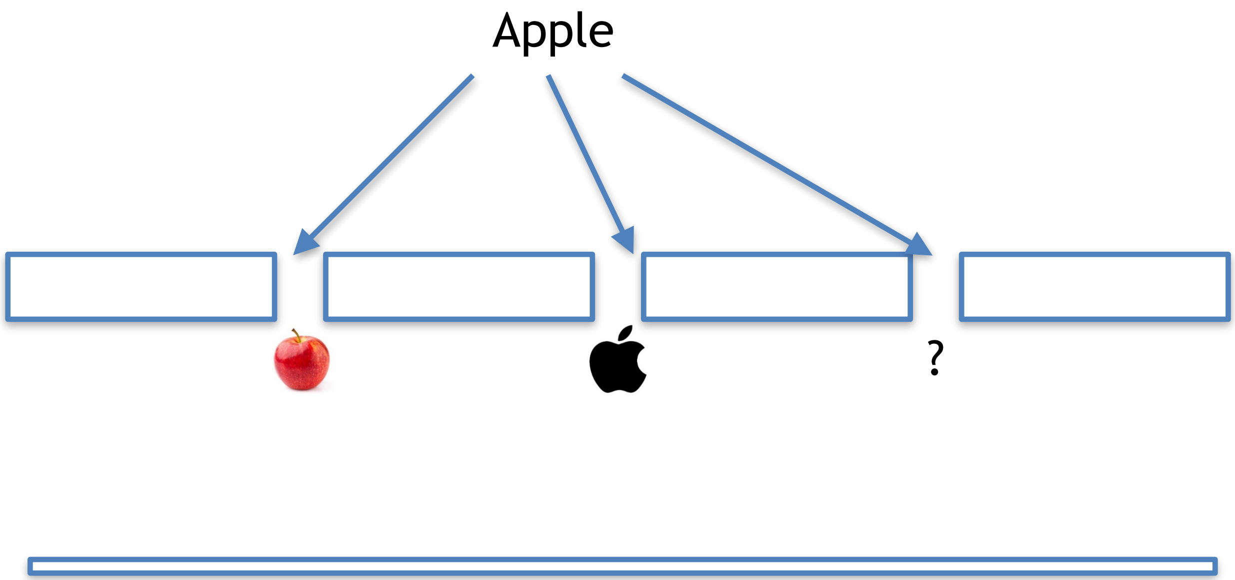
1. What is the concrete meaning of a single neutron? And how does it work? (*probability*)
2. What did we learning after training? (*unifying all the subcomponents in a single space and therefore they can mutually interpret each other*)

Inspiration

- Distributed representation
- Understanding a single neutron
- How it works
- What it learned

How to understand distributed representation (word vector)?

Distributed representation vs Superposition state over sememes



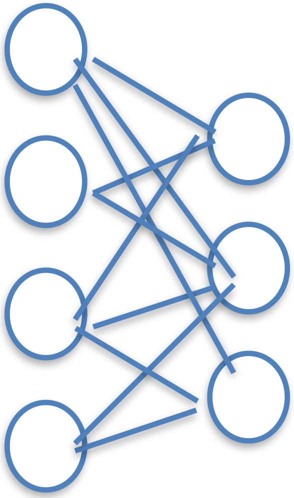
|Apple> = a | > + b | > + … c |?>

Decomposing word vector

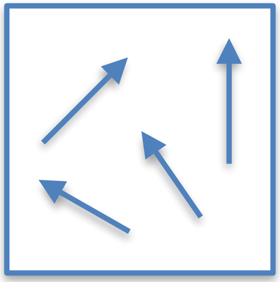
- Amplitude vector (unit-length vector)
    - Corresponding weight for each sememe
- Phase vector (each element ranges [0,2Pi])
    - Higher-lever semantic aspect
- Norm
    - Explicit weight for each word.

How to understand the value of a single neutron?

One neutron or a group of neutrons?



Vanilla neutrons



Capsules by Hinton

How does the neural network work?

Driven by probability?

Classical probability: **set**-based probability theory
(countable) events are in a discrete space

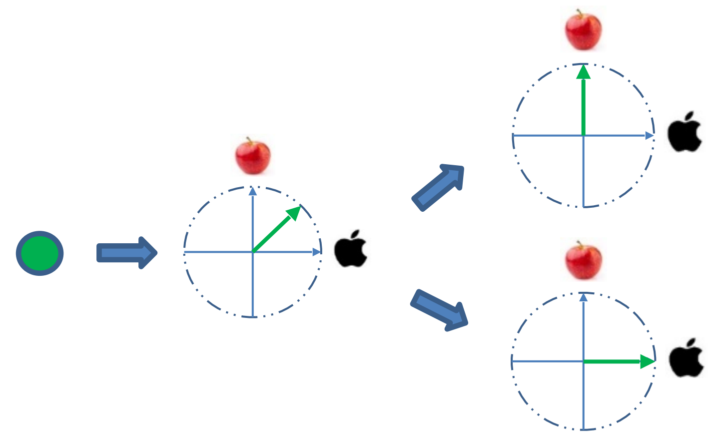
Quantum probability: **projective geometry** based theory
(uncountable) events are in a continuous space

Uncertainty in Language/QT

- A **single word** may have multiple meanings

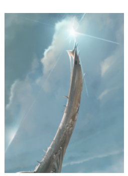
“Apple”

- Uncertainty of a pure state



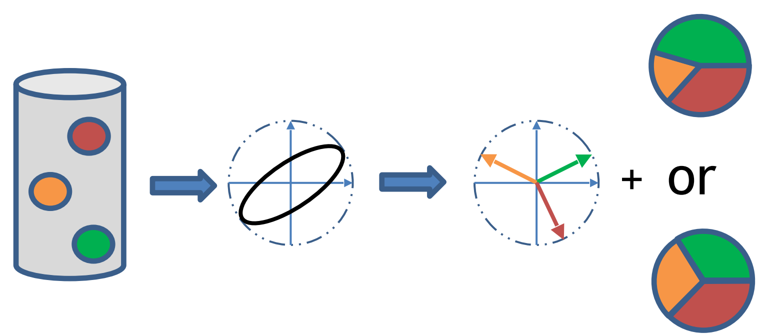
- **Multiple words** may be combined in different ways

“Ivory tower”



UNIVERSITY OF MINNESOTA

- Uncertainty of a mixed state

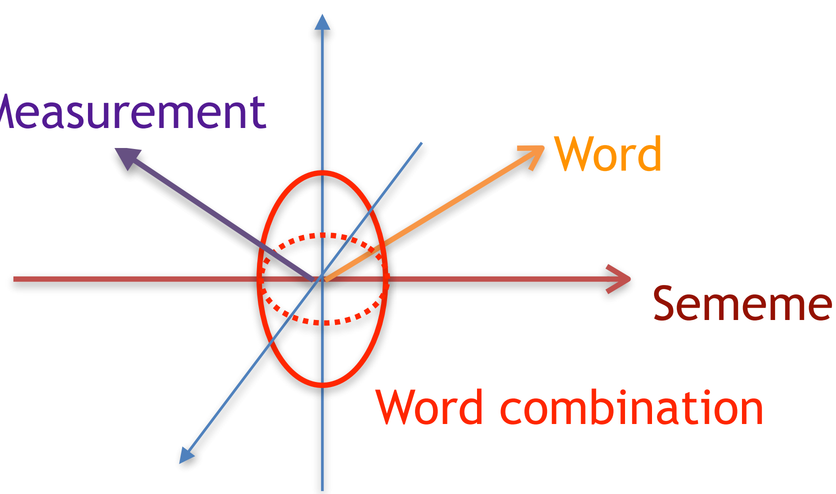


What did it learn?

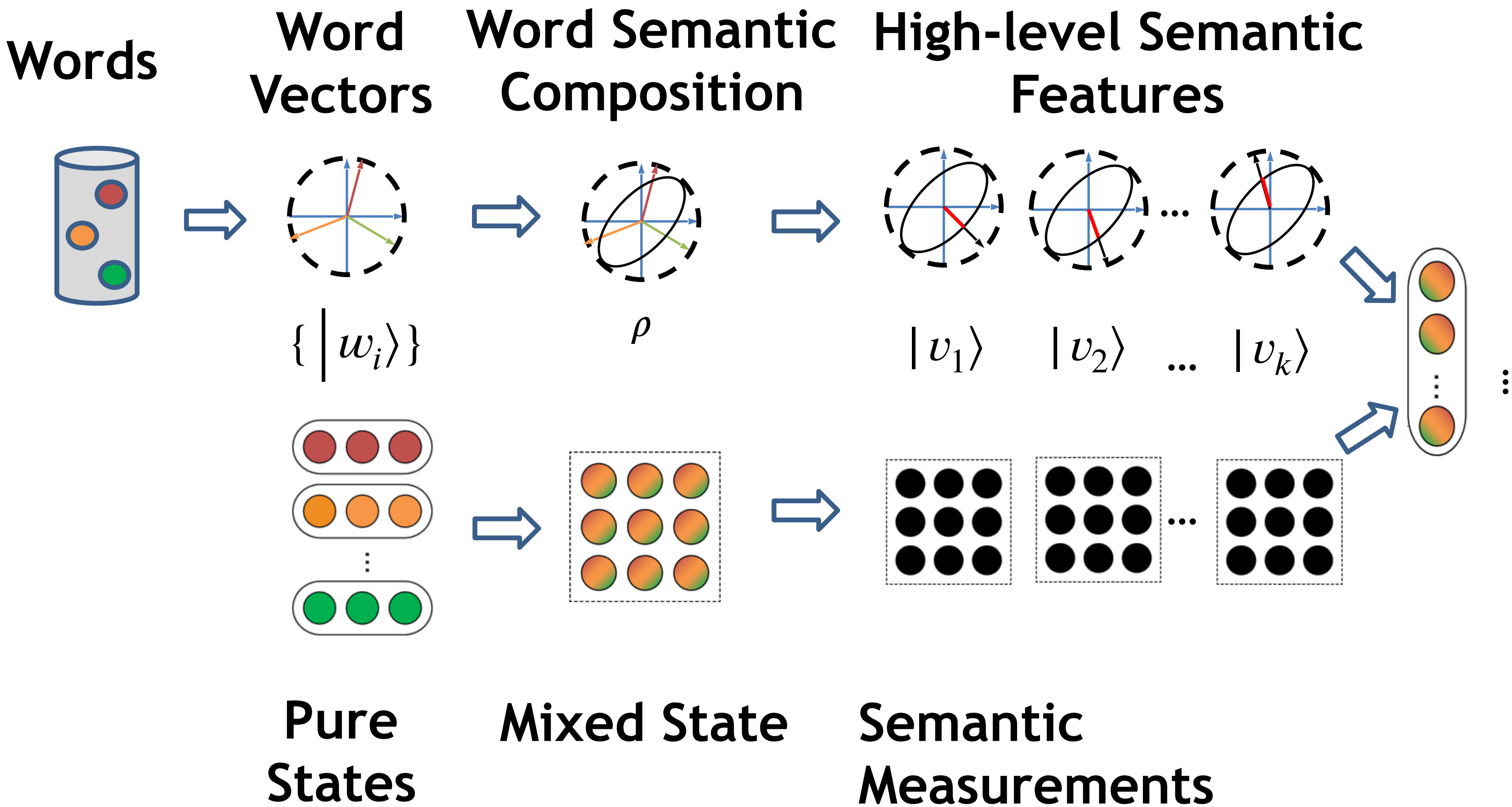
How to interpret trainable components?

- A unified quantum view of different levels of linguistic units
  - Sememes
  - Words
  - Word Combinations
  - Sentences

Therefore they can mutually interpret each other

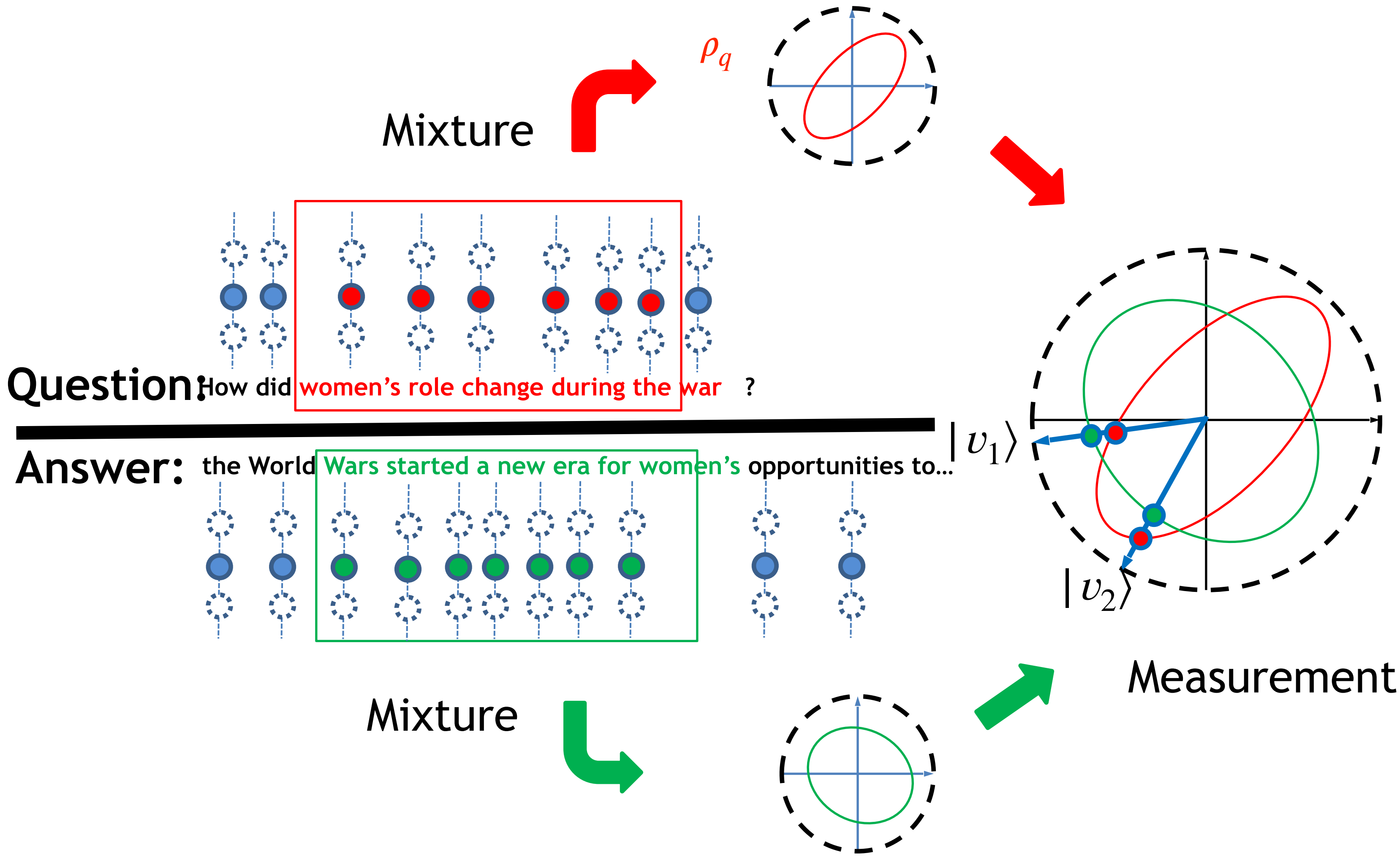


Semantic Hilbert Space

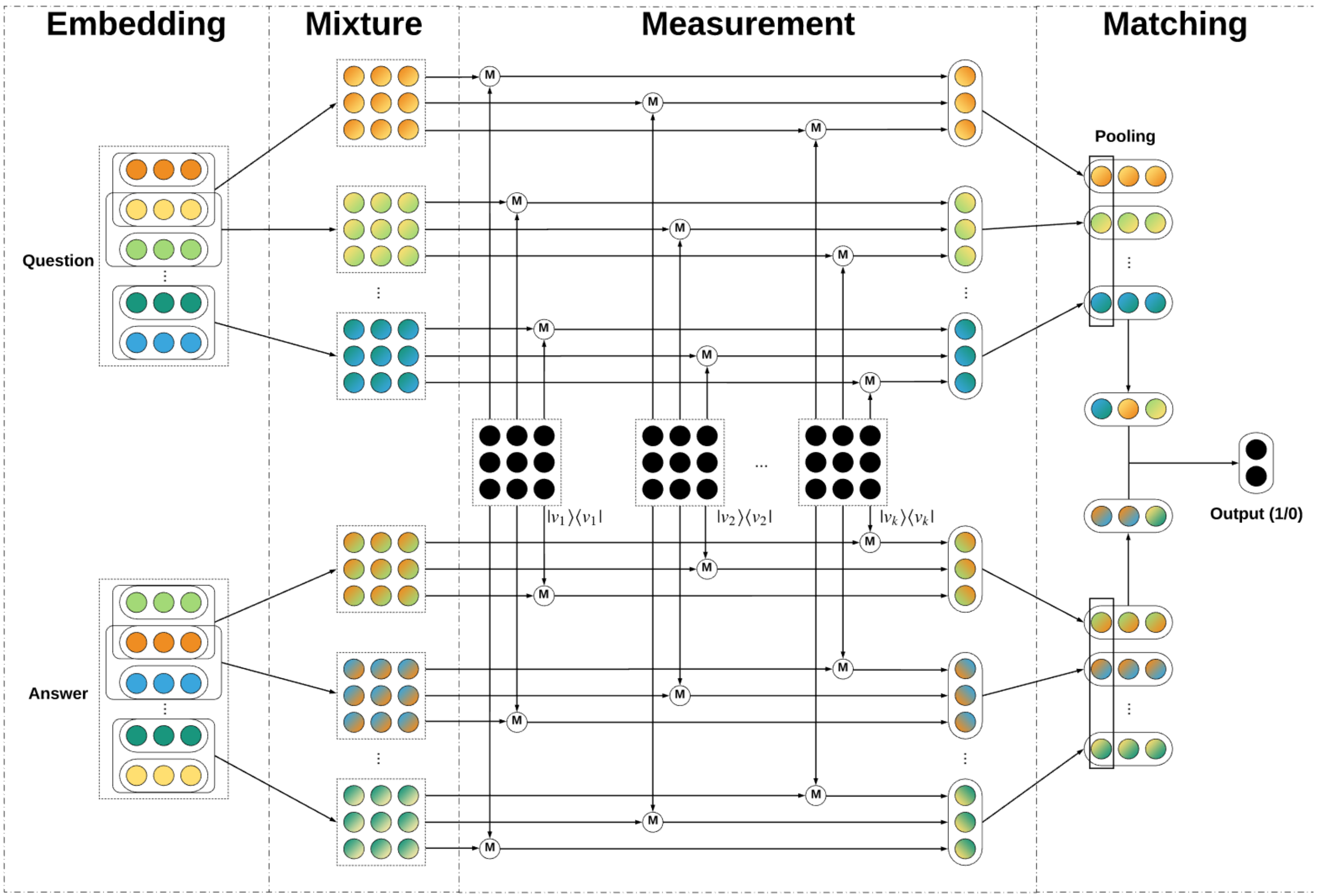


Benyou Wang*, Qiuchi Li*, Massimo Melucci, and Dawei Song. *Semantic Hilbert Space for Text Representation Learning*. In *WWW2019*

Application to Text Matching



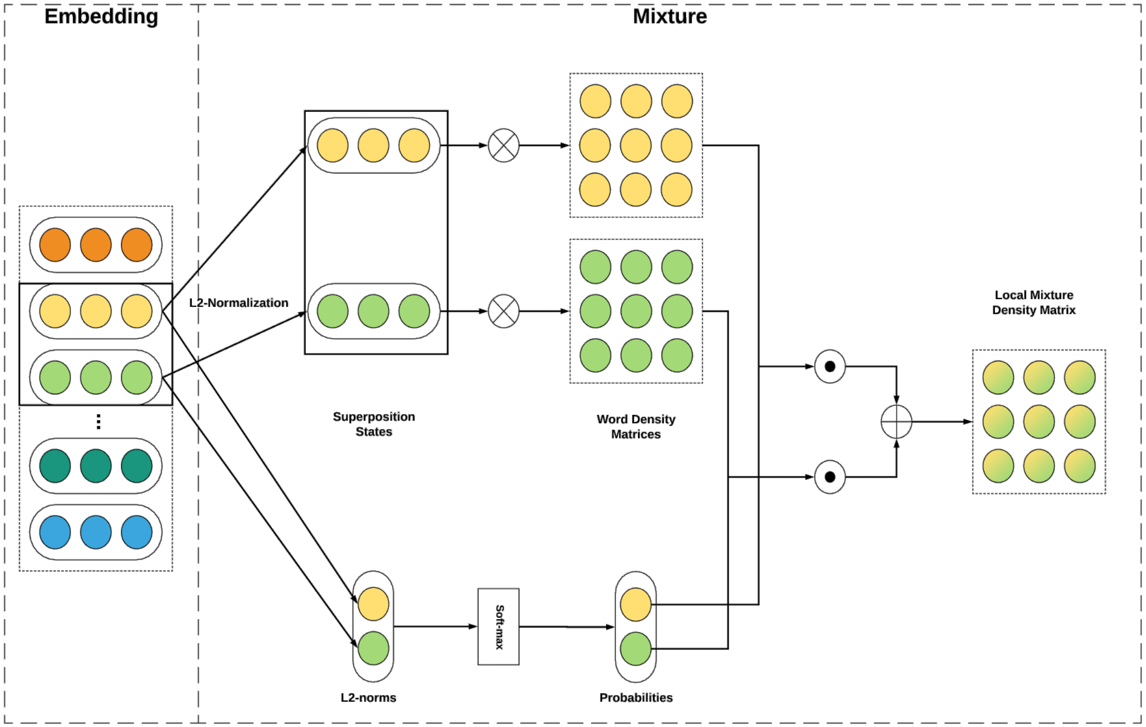
Complex-valued Network for Matching



X

- L2-normed word vectors as superposition states

- Softmax-normalized word L2-norms as mixture weights



Experiment Result

- Effectiveness
  - Competitive compared to strong baselines
  - Outperforms existing quantum-inspired QA model (Zhang et al. 2018)

Model	MAP	MRR
Bigram-CNN	0.5476	0.6437
LSTM-3L-BM25	0.7134	0.7913
LSTM-CNN-attn	0.7279	0.8322
aNMM	0.7495	0.8109
MP-CNN	**0.7770**	0.8360
CNTN	0.7278	0.7831
PWIM	0.7588	0.8219
QLM	0.6780	0.7260
NNQLM-I	0.6791	0.7529
NNQLM-II	0.7589	0.8254
CNM	0.7701	**0.8591**
Over NNQLM-II	1.48% ↑	4.08% ↑

Experiment Results on TREC QA Dataset. The best performing values are in bold.

Model	MAP	MRR
Bigram-CNN	0.6190	0.6281
QA-BILSTM	0.6557	0.6695
AP-BILSTM	0.6705	0.6842
LSTM-attn	0.6639	0.6828
CNN-Cnt	0.6520	0.6652
QLM	0.5120	0.5150
NNQLM-I	0.5462	0.5574
NNQLM-II	0.6496	0.6594
CNM	**0.6748**	**0.6864**
Over NNQLM-II	3.88% ↑	4.09% ↑

Experiment Results on WikiQA Dataset.The best performing values are in bold.

Experiment Result

- Ablation Test
  - ✓ Complex-valued Embedding
    - non-linear combination of amplitudes and phases
  - ✓ Local Mixture Strategy
  - ✓ Trainable Measurement

Setting	MAP	MRR
FastText-MaxPool	0.6659 (0.1042↓)	0.7152 (0.1439↓)
CNM-Real	0.7112 (0.0589↓)	0.7922 (0.0659↓)
CNM-Global-Mixture	0.6968 (0.0733↓)	0.7829 (0.0762↓)
CNM-trace-inner-product	0.6952 (0.0749↓)	0.7688 (0.0903↓)
CNM	0.7701	0.8591

Transparency

Components	DNN	CNM
Sememe	-	basis one-hot vector / **basis state** $\{e \| e \in \mathcal{R}^n, \|\|e\|\|_2 = 1\}$ complete &orthogonal
Word	real vector $(-\infty, \infty)$	unit complex vector / **superposition state** $\{w \| w \in \mathcal{C}^n, \|\|w\|\|_2 = 1\}$
N-gram/ Word combinations	real vector $(-\infty, \infty)$	density matrix / **mixed system** $\{\rho \| \rho = \rho^*, tr(\rho) = 1$
Abstraction	CNN/RNN $(-\infty, \infty)$	projector / **measurement** $\{vv^T \| v \in \mathcal{C}^n, \|\|v\|\|_2 = 1\}$
Sentence representation	real vector $(-\infty, \infty)$	real value/ **measured probability** $(0, 1)$

Physical meanings and constraints.

- With well-constraint complex values, CNM components can be **explained** as concrete **quantum states** at design phase

Post-hoc Explainability

- Visualisation of word weights and matching patterns

Q: Who is the [president or chief executive of Amtrak] ?
A: ...said George Warrington, [Amtrak 's president and chief executive].

Q: How did [women 's role change during the war] ?
A: the [World Wars started a new era for women 's] opportunities to...

Semantic Measurements

- Each measurement is a pure state
- Understand measurement via neighboring words

	Selected neighborhood words for a measurement vector
1	andes, nagoya, inter-american, low-caste
2	cools, injection, boiling, adrift
3	andrews, paul, manson, bair
4	historically, 19th-century, genetic, hatchback
5	missile, exile, rebellion, darkness

Selected learned measurements for TREC QA. They were selected according to nearest words for a measurement vector in Semantic Hilbert Space.

Word L2-norms

- Rank words by l2-norms and select most important and unimportant words

	Selected words
Important	studio, president, women, philosophy scandinavian, washingtonian, berliner, championship defiance, reporting, adjusted, jarred
Unimportant	71.2, 5.5, 4m, 296036, 3.5 may, be, all, born movements, economists, revenues, computers

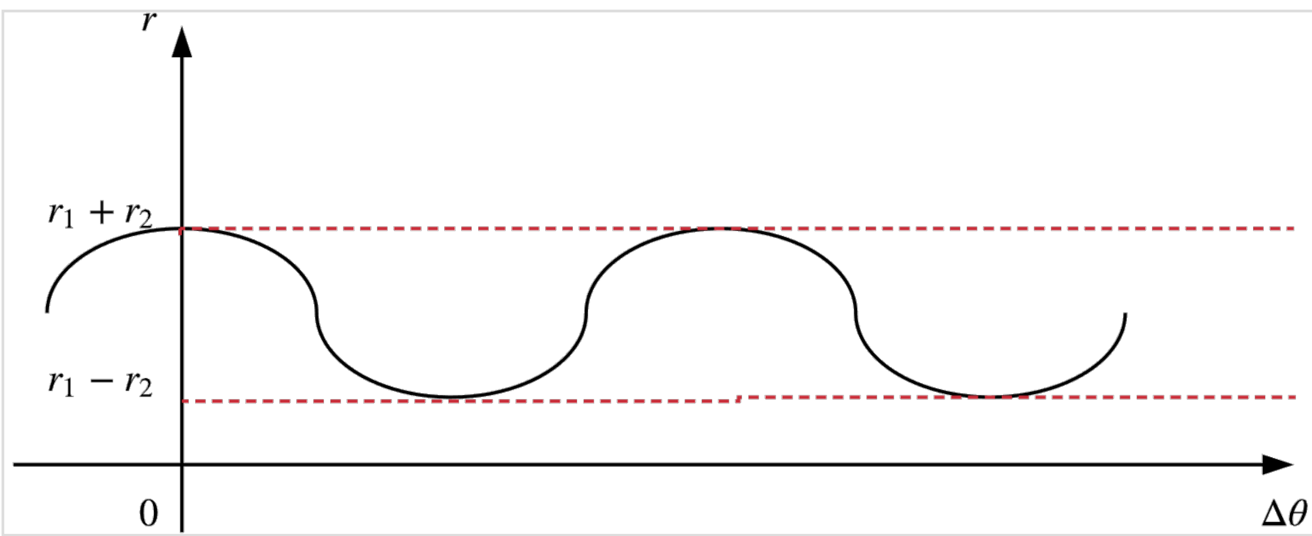
Selected learned important words in TREC QA. All words are converted to lower cases.

Complex-valued Embedding

- Non-linearity

Ivory tower ≠ Ivory + tower

- Gated Addition



Other Potentials

- Interpretability
- Robustnees
  - orthogonal projection subspaces (measurements)
- Transferness
  - Selecting some measurements is a kind of sampling. More measurements, in principle, lead to more accurate inference with respect to the given input. (like ensemble strategy)
  - Reusing the trained measurement from one dataset to another dataset makes sense, especially that recent works tends to use a given pertained language model to build the input features

Conclusion & Future Work

- Conclusion
  - Interpretability for language understanding
  - Quantum-inspired complex-valued network for matching
    - Transparent & Post-hoc Explainable
    - Comparable to strong baselines
- Future Work
  - Incorporation of state-of-the-art neural networks
  - Experiment on larger datasets

- Contacts
  - qiuchili@dei.unipd.it
  - wang@dei.unipd.it

- Source Code
  - github.com/wabyking/qnn

Contents

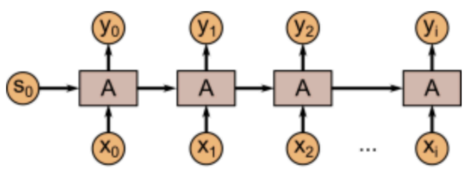
- History of quantum inspired IR/NLP
- Basics from Quantum theory
- Semantic Hilbert Space—NAACL best paper
- **Future work with Quantum Theory**

What the companies care

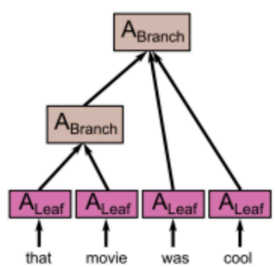
- How to improve SOTA
  - Performance
  - Efficiency
  - Interpretability

Position and order

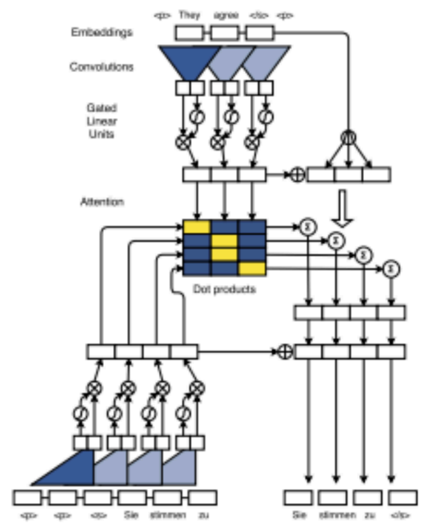
- Transformer without position embedding is position-insensitive



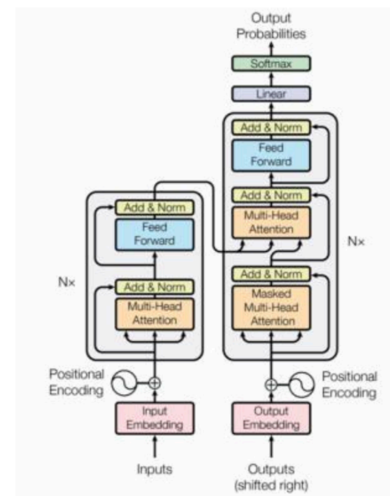
Recurrent Neural Net



Recursive Neural Net

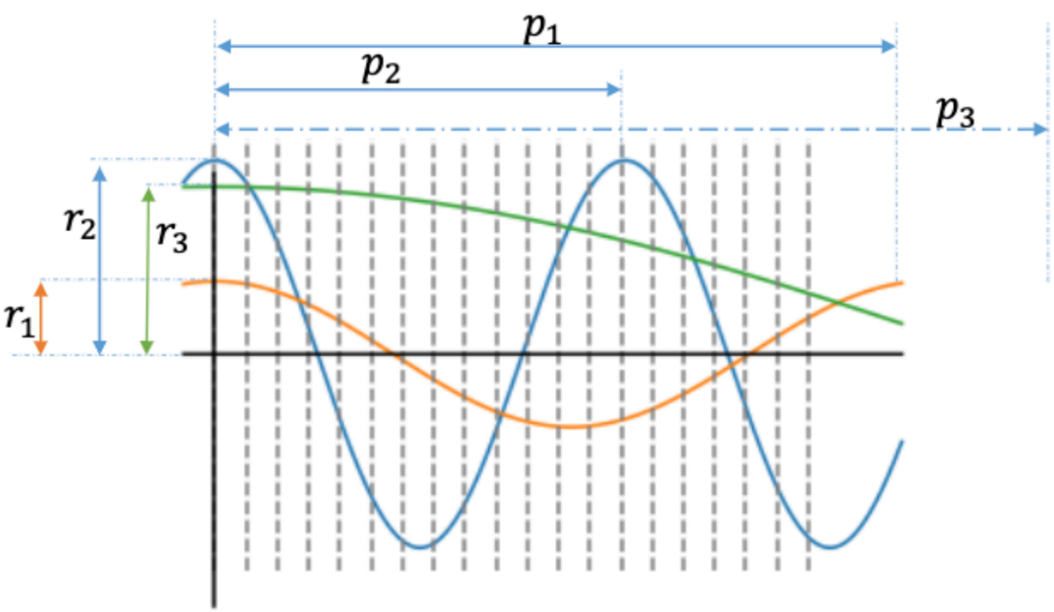


Conv seq2seq



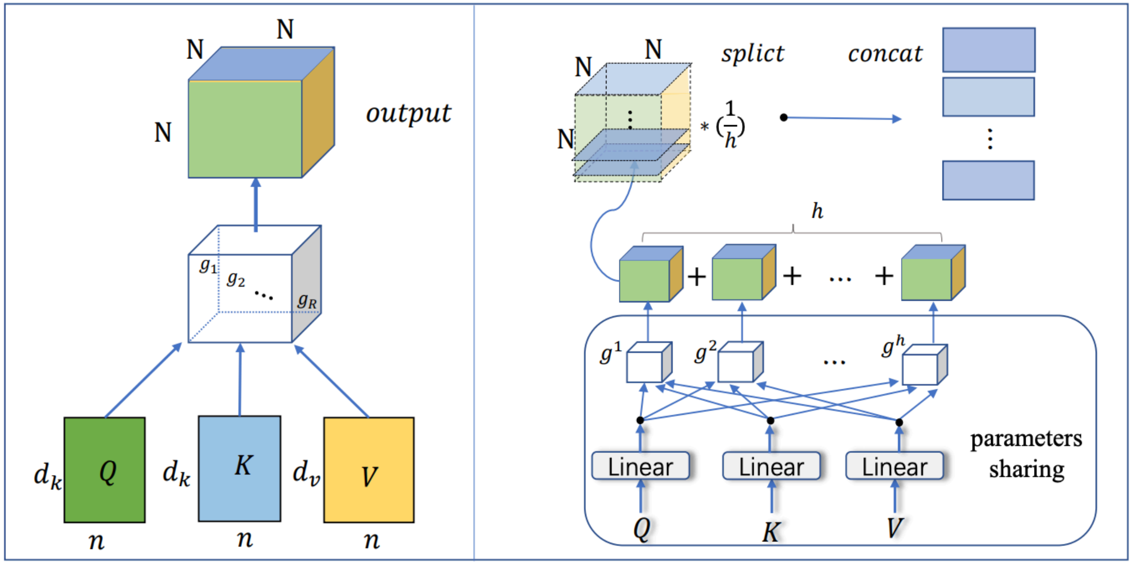
Transformer

Encoding order in complex-valued embedding



Efficiency

- Tensor decomposition



Ma et.al. A Tensorized Transformer for Language Modeling, https://arxiv.org/pdf/1906.09777.pdf

Efficiency

Results in language model

Model	PTB			WikiText-103		
	Params	Val PPL	Test PPL	Params	Val PPL	Test PPL
LSTM+augmented loss [15]	24M	75.7	48.7	–	–	48.7
Variational RHN [38]	23M	67.9	65.4	–	–	45.2
4-layer QRNN [21]	–	–	–	151M	–	33.0
AWD-LSTM-MoS [36]	22M	58.08	55.97	–	29.0	29.2
Transformer+adaptive input [1]	24M	59.1	57	247M	19.8	20.5
Transformer-XL [7]	24M	56.72	54.52	151M	23.1	24.0
Transformer-XL+TT [18]	18 M	57.9*	55.4*	130M	23.61*	25.70*
Tensorized Transformer core-1	12M	60.5	57.9	80.5M	22.7	20.9
Tensorized Transformer core-2	12M	**54.25**	**49.8**	86.5M	**19.7**	**18.9**

Results and compression with Transformer on WMT-16 English-to-German translation.

Model	Params	BLEU
Base-line [30]	–	26.8
Linguistic Input Featurec [29]	–	28.4
Attentional encoder-decoder + BPE [30]	–	34.2
Transformer [34]	52M	34.5*
Tensorized Transformer core-1	21M	34.10
Tensorized Transformer core-2	21.2M	**34.91**

Ma et.al. A Tensorized Transformer for Language Modeling, https://arxiv.org/pdf/1906.09777.pdf

Interpretability

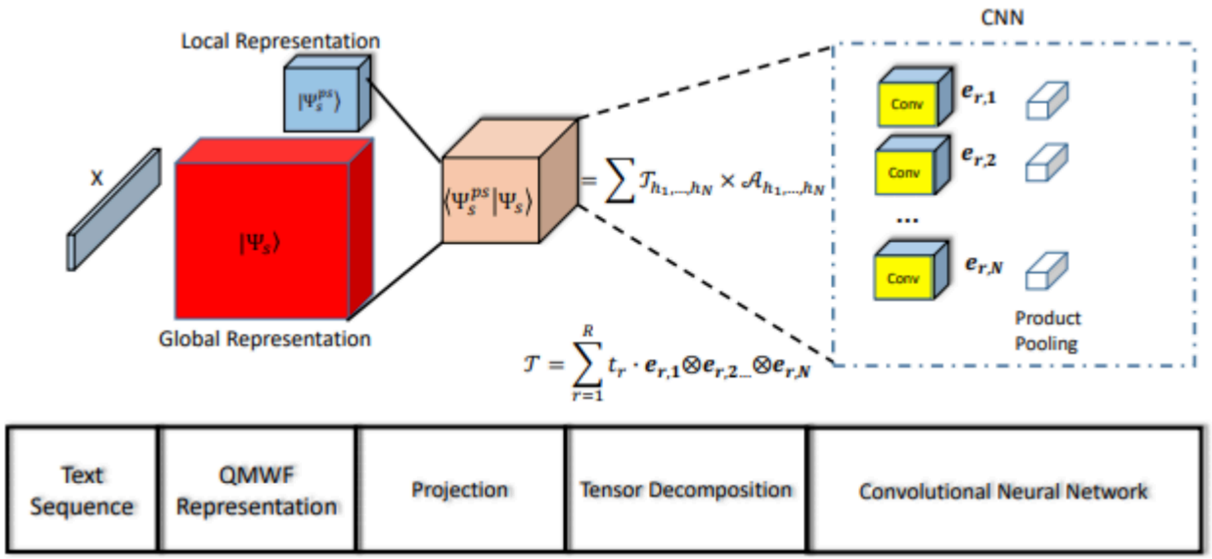
with the connection between tensor and NN

Correspondence between languages of Tensor Analysis and Deep Learning.

Tensor Decompositions	**Deep Learning**
CP-decomposition	shallow network
TT-decomposition	RNN
HT-decomposition	CNN
rank of the decomposition	width of the network

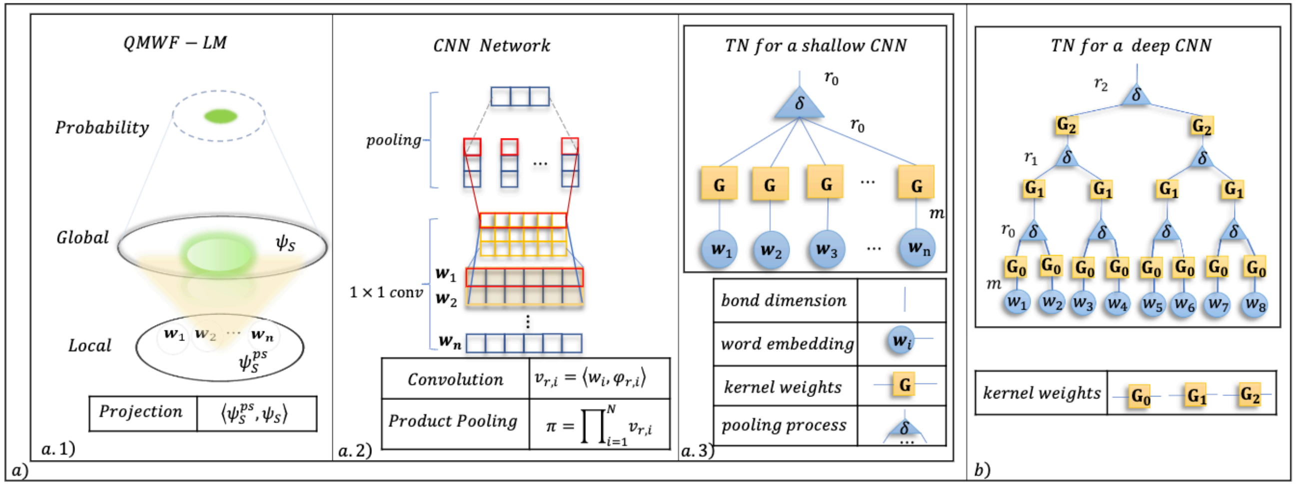
Khrulkov, Valentin, Alexander Novikov, and Ivan Oseledets. "Expressive power of recurrent neural networks." *arXiv preprint arXiv:1711.00811* (2017). ICLR 2018

Interpretability (1):
Tensor Vs DL



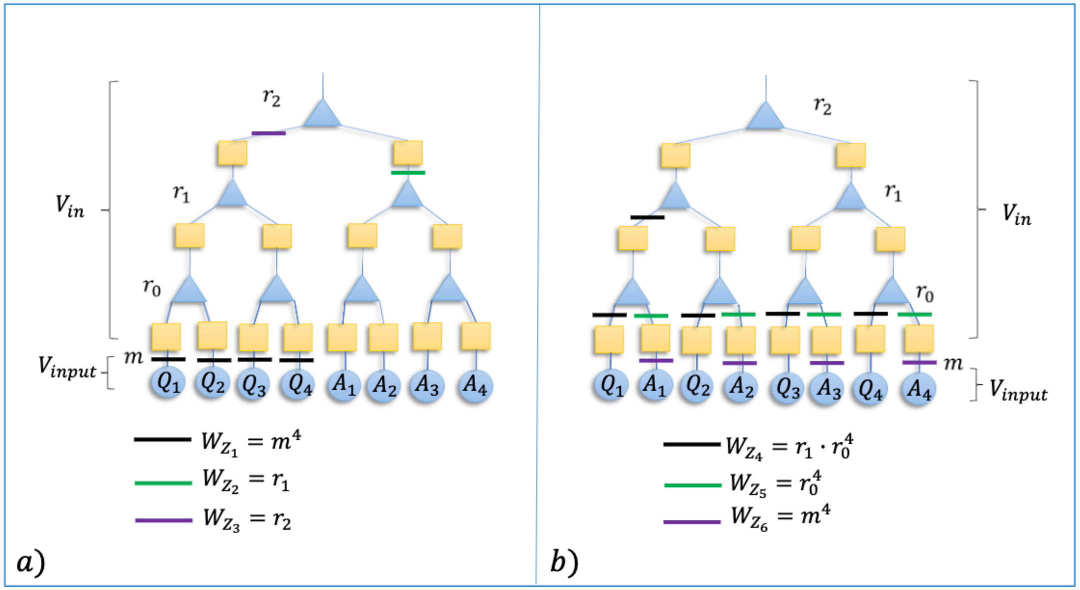
Zhang P, Su Z, Zhang L, **Wang B**, Song D. A quantum many-body wave function inspired language modeling approach. In Proceedings of the 27th ACM CIKM 2018 Oct 17 (pp. 1303-1312). ACM.

Interpretability (2):
Long-term and short-term dependency



Interpretability (2):

Long-term and short-term dependency



References

- Li, Qiuchi, Sagar Uprety, Benyou Wang, and Dawei Song. “Quantum-Inspired Complex Word Embedding.” *Proceedings of The Third Workshop on Representation Learning for NLP*, 2018, 50–57.
- Wang, Benyou, Qiuchi Li, Massimo Melucci, and Dawei Song. “Semantic Hilbert Space for Text Representation Learning.” In *The World Wide Web Conference*, 3293–3299. WWW ’19. New York, NY, USA: ACM, 2019.
- Lipton, Zachary C. “The Mythos of Model Interpretability.” *Queue* 16, no. 3 (June 2018): 30:31–30:57.
- Peter D Bruza, Kirsty Kitto, Douglas McEvoy, and Cathy McEvoy. 2008. “Entangling words and meaning”. *pages QI*, 118–124. College Publications.
- Bruza, P.D., and R.J. Cole. “Quantum Logic of Semantic Space: An Exploratory Investigation of Context Effects in Practical Reasoning.” In *We Will Show Them! Essays in Honour of Dov Gabbay*, 1:339–362. London, UK: College Publications, 2005.
- Diederik Aerts and Marek Czachor. “Quantum Aspects of Semantic Analysis and Symbolic Artificial Intelligence.” *Journal of Physics A: Mathematical and General* 37, no. 12 (2004): L123.
- Diederik Aerts and L. Gabora. “A Theory of Concepts and Their Combinations: II.: A Hilbert Space Representation.” *Kibernetes* 34, no. 1 (2004): 192–221.
- Diederik Aerts and L. Gabora. “A Theory of Concepts and Their Combinations I: The Structure of the Sets of Contexts and Properties.” *Kybernetes* 34 (2004): 167–191.
- Diederik Aerts and Sandro Sozzo. “Quantum Entanglement in Concept Combinations.” *International Journal of Theoretical Physics* 53, no. 10 (October 1, 2014): 3587–3603.
- Sordoni, Alessandro, Jian-Yun Nie, and Yoshua Bengio. “Modeling Term Dependencies with Quantum Language Models for IR.” In *Proceedings of the 36th International ACM SIGIR Conference on Research and Development in Information Retrieval*, 653–662. SIGIR ’13. New York, NY, USA: ACM, 2013.

Thanks!

Related Works

- Quantum-inspired Information Retrieval (IR) models
  - Query Relevance Judgement (QPRP, QMR,…)
  - Quantum Language Model (QLM) and QLM variants

- Quantum-inspired NLP models
  - Quantum-theoretic approach to distributional semantics (Blacoe et al. 2013; Blacoe 2015a; Blacoe 2015b)
  - NNQLM (Zhang et al. 2018)
  - Quantum Many-body Wave Function (QMWF) (Zhang et al. 2018)
  - Tensor Space Language Model (TSLM) (Zhang et al. 2019)
  - QPDN (Li et al. 2018; Wang et al. 2019)