DEPARTMENT OF INFORMATION ENGINEERING
UNIVERSITY OF PADOVA

QUARTZ
Quantum Information Access and Retrieval Theory

Quantum formulations for language:
understand words as particles

Benyou Wang

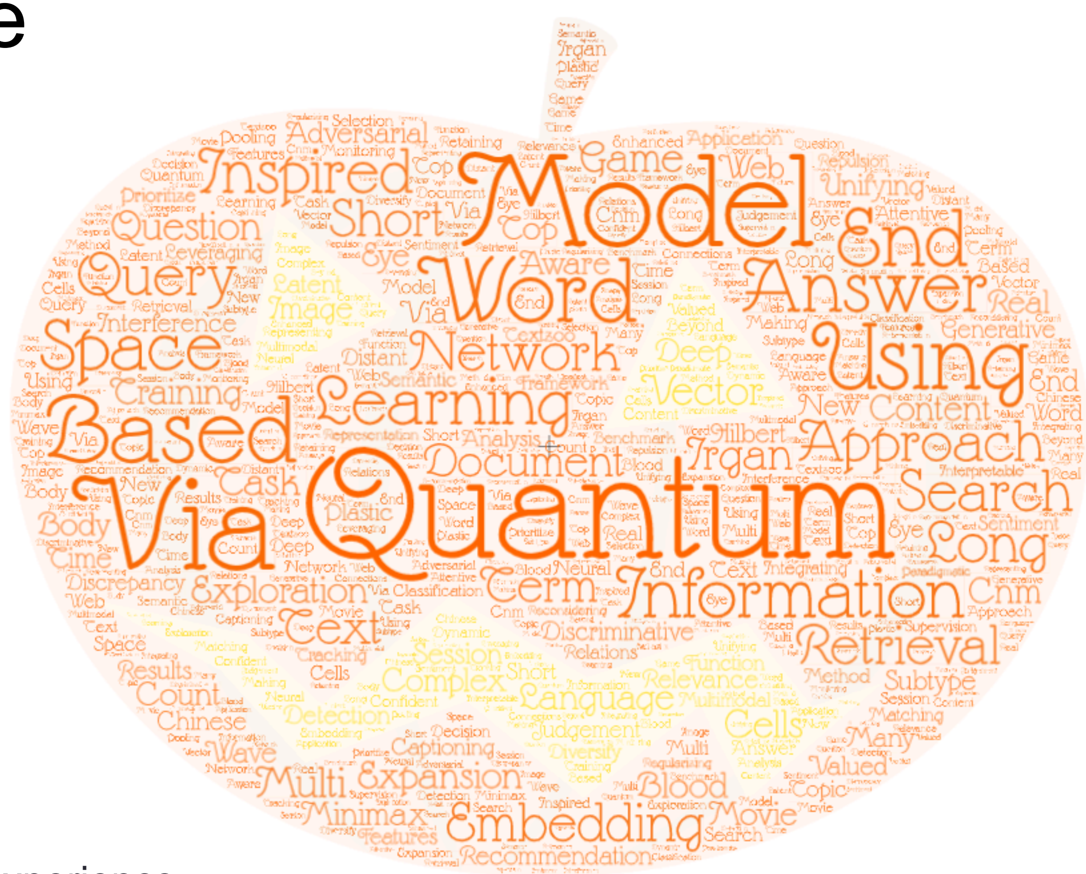
Supervised by Massimo Melucci and Emanuele Di Buccio

University of Padua

Amsterdam, Netherlands, 25/10/2019

About me

Benyou Wang

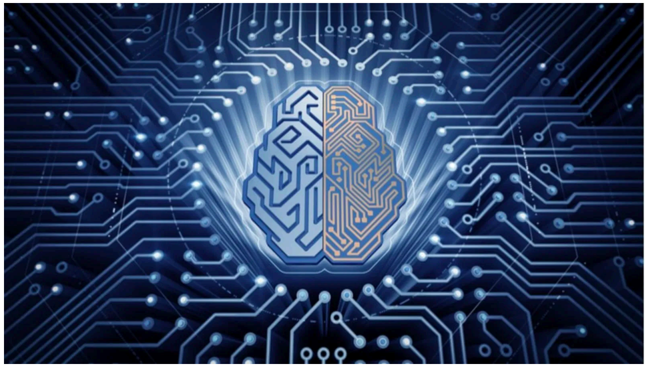


- Research Experience
  - 2014-2017 Master student, Tianjin University, China
  - 2017-2018 Associate Researcher, Tencent
  - 2018-2021 Marie Curie Researcher and PhD student, University of Padua
- Awards
  - SIGIR 2017 Best paper Honourable mention
  - NAACL 2019 Best Explainable NLP Paper

Quantum theory **outside** Physics

Using quantum ways to process information

- Quantum computing
  - [Michael A. Nielsen, Isaac L. Chuang. 2011. Quantum Computation and Quantum Information, 10th edition. Cambridge University Press]
  - Arute .et.al. Quantum supremacy using a programmable superconducting processor. Nature. 23 October 2019.
- Social science and cognition science
  - [Jerome R. Busemeyer and Peter D. Bruza. 2013. Quantum Models of Cognition and Decision. Cambridge University Press]
  - [E. Haven and A. Khrennikov. 2013. Quantum Social Science. Cambridge University Press.]
- **Information retrieval**
  - [Van Rijsbergen. 2004. The geometry of information retrieval. Cambridge University Press.]
  - [Massimo Melucci. 2016. Introduction to information retrieval and quantum mechanics. Springer Berlin Heidelberg.]



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

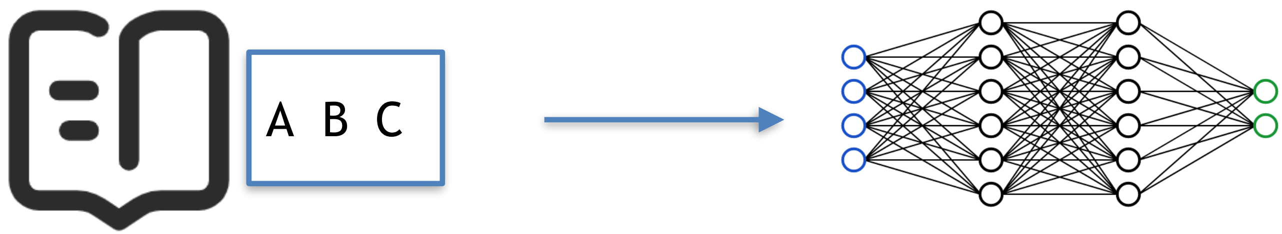
- Quantum IR does not rely on quantum computing/cognition, but share the same mathematical foundation to **probabilistically** describe the world

Contents

- Background
- Semantic Hilbert Space [1]
  - **Why quantum?**
  - How does it work?
  - Any benefits?
- Possible Future works

[1] Li, Qiuchi*, **Benyou Wang***, and Massimo Melucci. "CNM: An Interpretable Complex-valued Network for Matching." NAACL 2019 Best Explainable NLP Paper.

Probability-driven for language



discrete

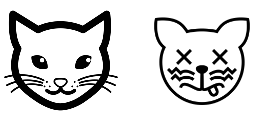
Vector spaces

Neural networks usually transform a discrete token to a vector.
We need a probability theory to describe uncertainty in vector spaces.

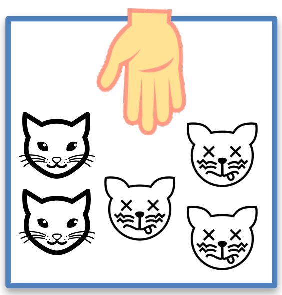
Quantum Probability Theory

a probability theory defining on **vector spaces**

Set-based Probability Theory



alive dead



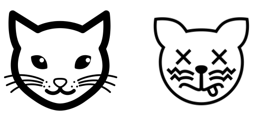
Q: Should the randomly-chosen cat dead or alive ?

A: 0.4 to be alive and 0.6 to be dead

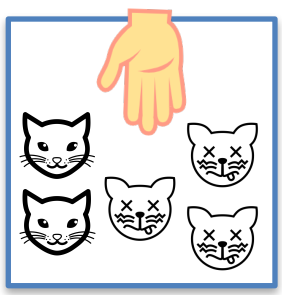
Quantum Probability Theory

a probability theory defining on **vector spaces**

Set-based Probability Theory



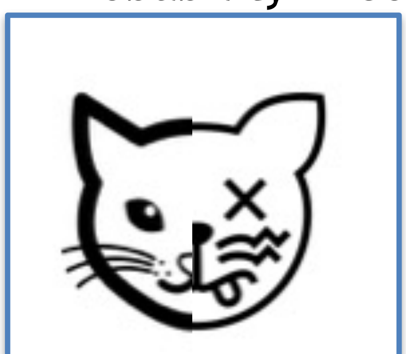
alive dead



Q: Should the randomly-chosen cat dead or alive ?

A: 0.4 to be alive and 0.6 to be dead

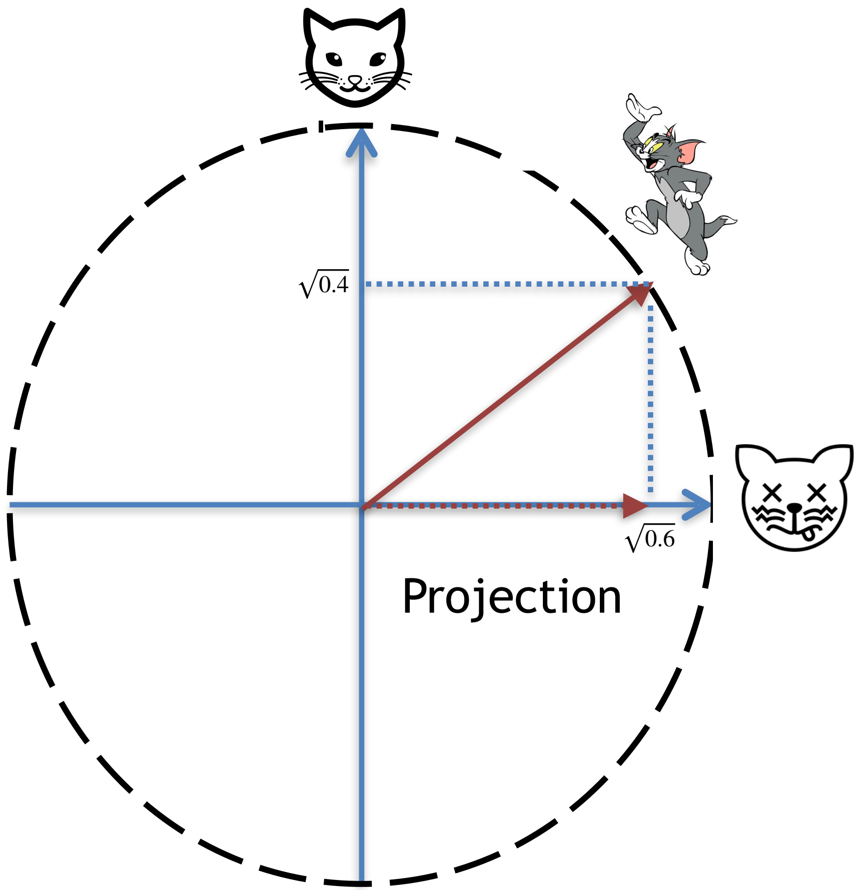
Quantum Probability Theory - **vector-based**



Q: Are these cat dead or alive?

A: 0.5.001 to be alive and 0.499 to be dead

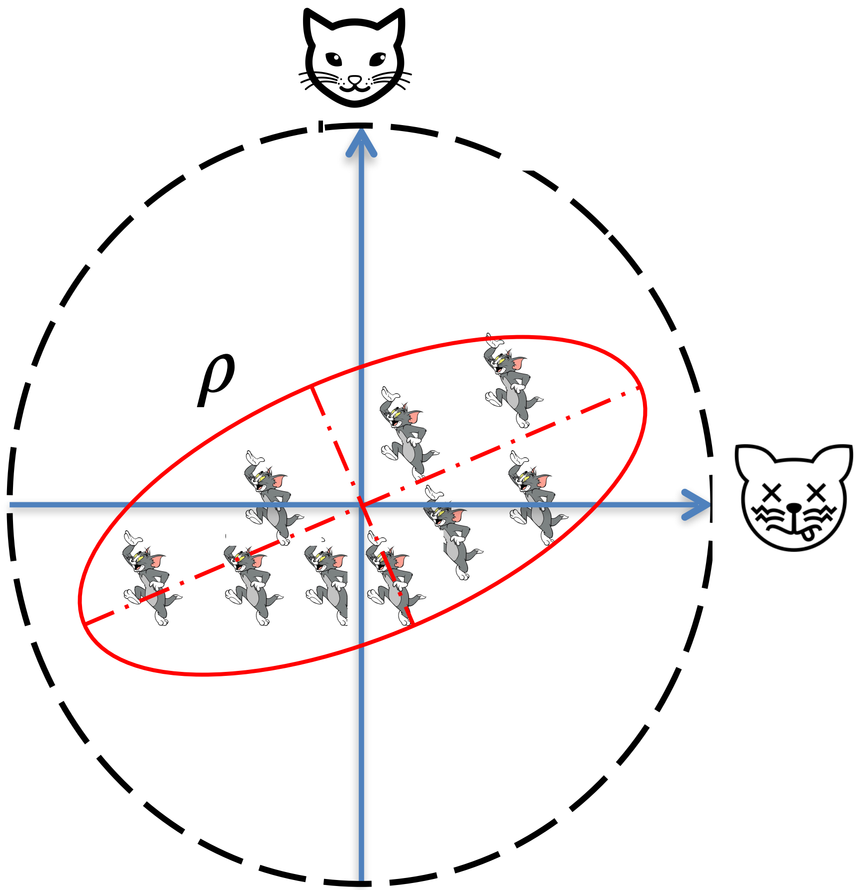
Probability theory in vector spaces
for single object



Square of the projection length denotes the probability

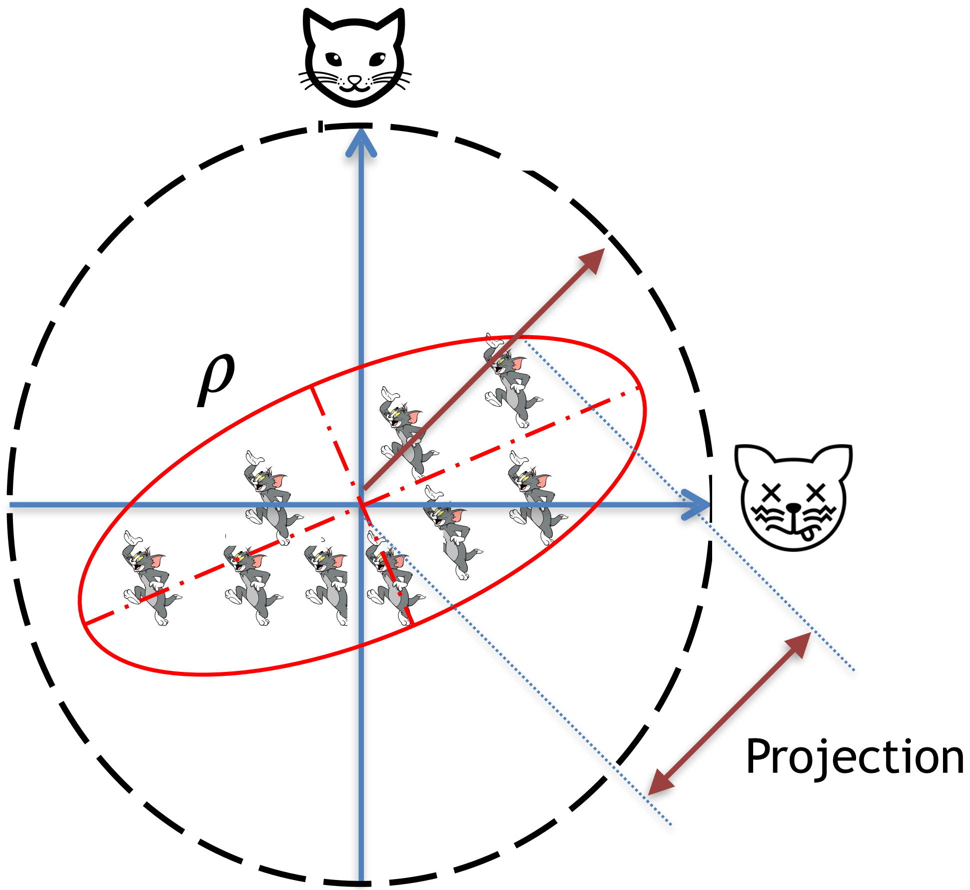
Probability theory in vector spaces

for many objects



Square of the projection length denotes the probability

Probability theory in vector spaces
for many objects



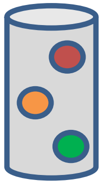
Square of the projection length denotes the probability

Contents

- Background
- Semantic Hilbert Space for matching
  - Why quantum?
  - **How does it work?**
  - Any benefits?
- Possible Future works

Semantic Hilbert Space

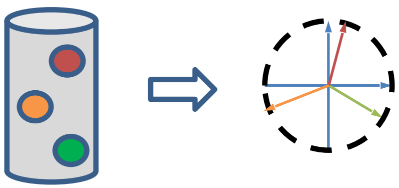
Words



a document is considered as a bag of words

Semantic Hilbert Space

Words **Word Vectors**

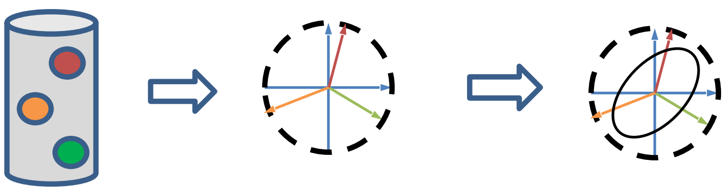


Look up its **complex-valued word embedding** for each word

*A **pure state** to describe a single particle*

Semantic Hilbert Space

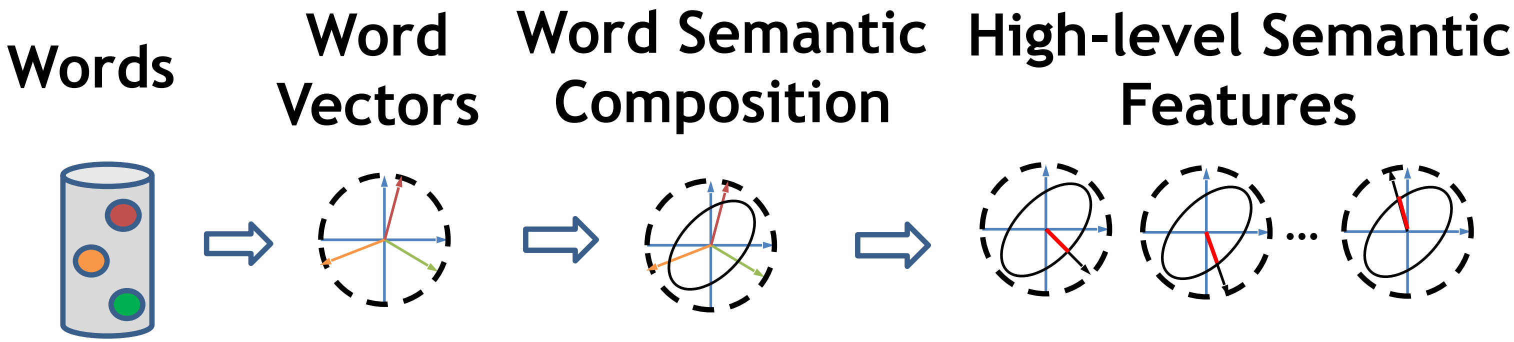
Words **Word Vectors** **Word Semantic Composition**



Build a **density distribution** for many words

*A **mixed state** to describe a system with many particles*

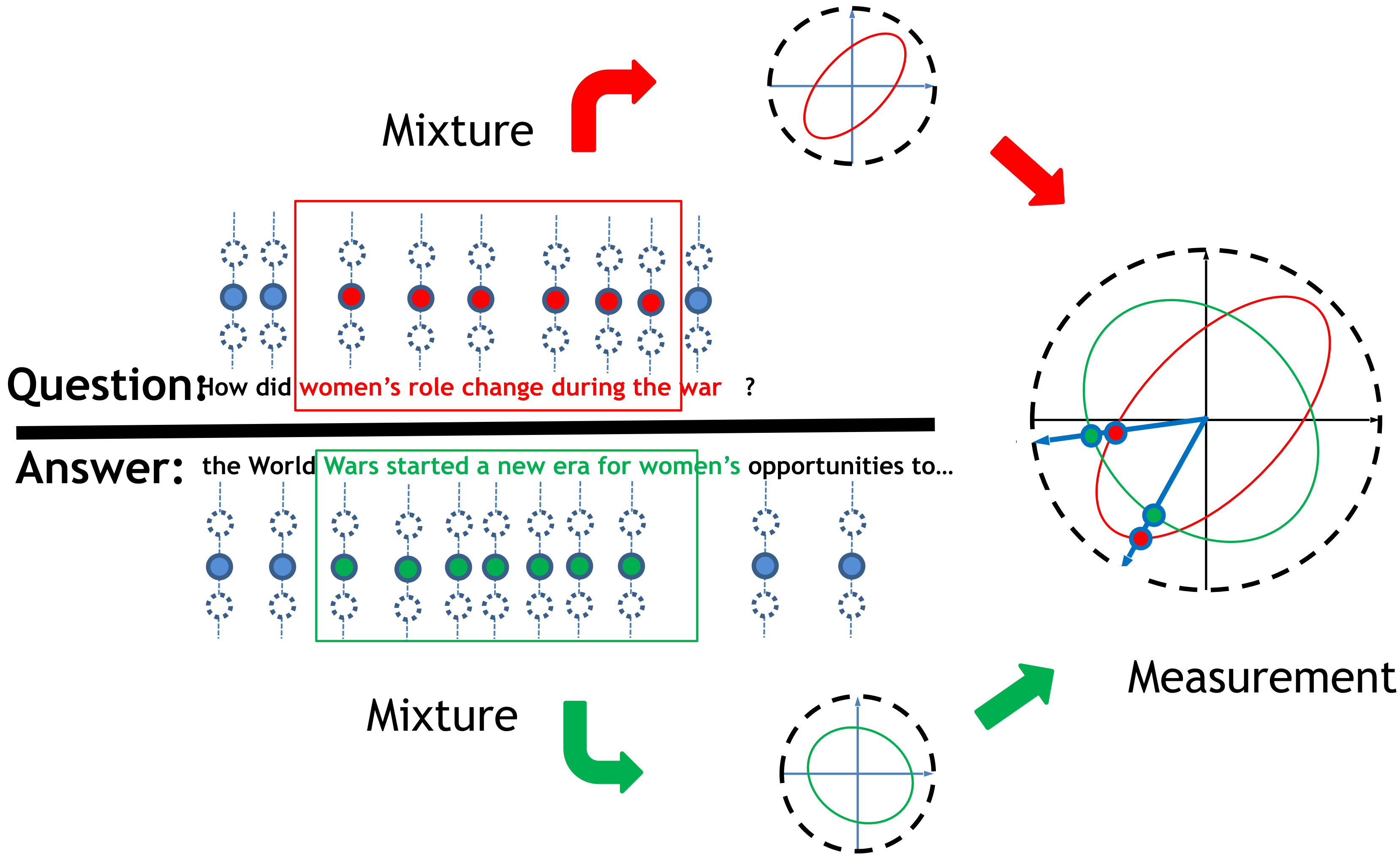
Semantic Hilbert Space



Use projections to know the document

A probability distribution based on a group of measurements

Application to text matching



Experiment result

- Effectiveness
  - Competitive compared to strong baselines like RNN, CNN
  - 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% ↑

Experimental results on WIKI QA dataset,

Code : github.com/wabyking/qnn.git

Contents

- Background
- Semantic Hilbert Space—NAACL best explainable NLP paper
    - Why quantum?
    - How does it work ?
    - **Any benefits?**
        - **understanding distributed representations via superposition**
        - **well-mathematically-constrained framework**
- Possible Future works

1) Distributed representation as Superposition state

Assume the smallest semantic unit (atomic or indivisible) as Sememe

Words are considered as one or a combination of **sememes**: like

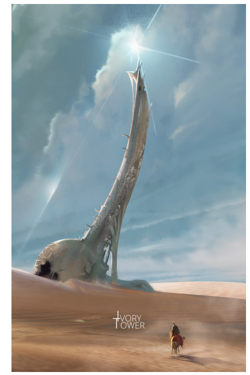
$$blacksmith = S_{human} \bigoplus S_{occupation} \bigoplus S_{metal} \bigoplus S_{industrial} \cdots$$

Assuming sememes are independent and orthogonal as one hot vectors.

$$\overrightarrow{balcksmith} = [e_1, e_2, e_3, e_4, \cdots]^T [S_{human}, S_{occupation}, S_{metal}, S_{industrial} \cdots]$$

1) Benefit from complex-valued weights

Ivory tower ≠ Ivory + tower
= Ivory + tower + Δ



UNIVERSITY OF AMSTERDAM

Complex numerical example:

$$z_1 = \alpha_1 + \beta_1 i = r_1 e^{i\theta_1}. \qquad ||z_1||_2 = r_1^2$$

$$z_2 = \alpha_2 + \beta_2 i = r_2 e^{i\theta_2}. \qquad ||z_2||_2 = r_2^2$$

$$z_1 + z_2 = r_1 e^{i\theta_1} + r_2 e^{i\theta_2}. \qquad ||z_1 + z_2||_2 = r_1^2 + r_2^2 + \boxed{2 r_1 r_2 \cos(\theta_1 - \theta_2)}$$

Interference term

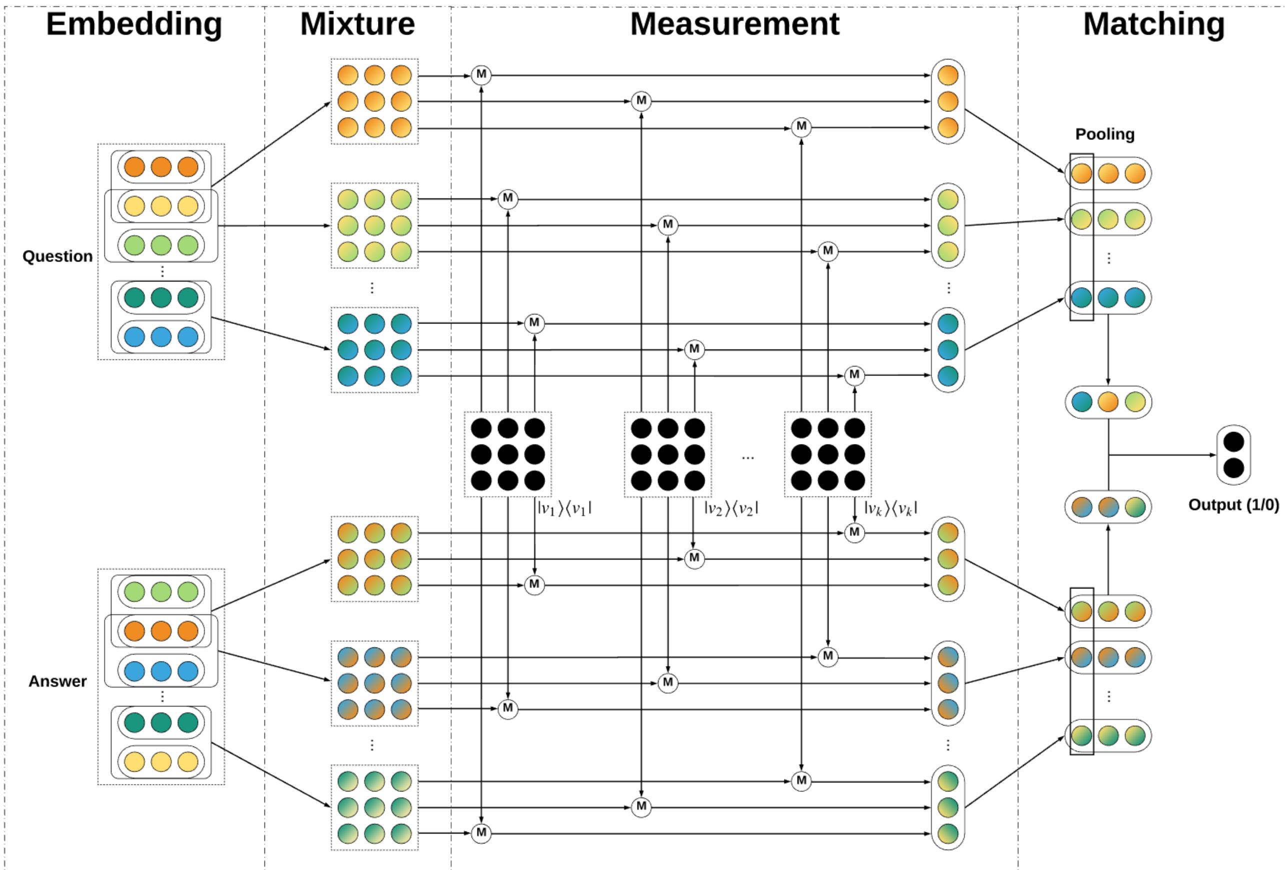
2) Transparency

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

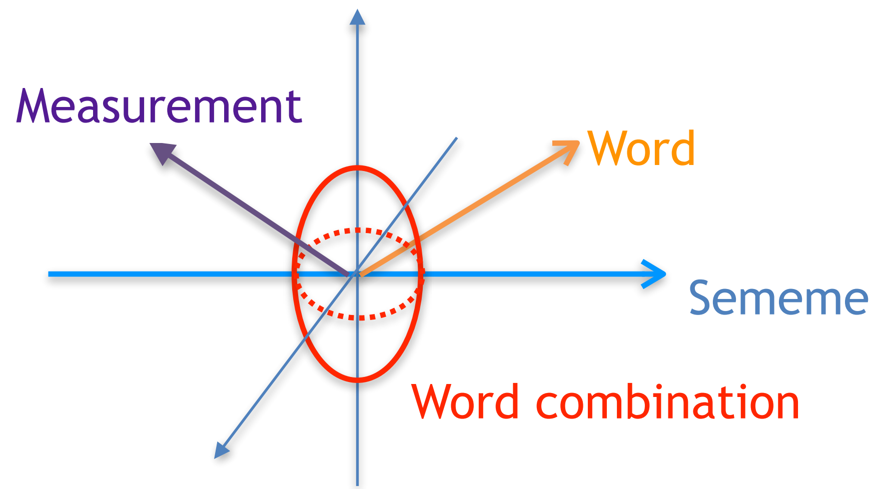
2) Meaning of each component



Embedding	Mixture	Measurement	Matching
Probability distribution for each vector - a word	Probability distribution for each matrix - a N-gram	Probability distribution for each measurement	Probability values

Most important learning components

2) A unified vector space



Understanding measurement via neighbouring 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.

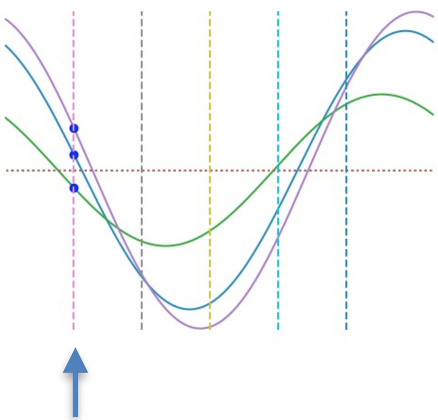
Van Rijsbergen. 2004. The geometry of information retrieval. Cambridge University Press.

Contents

- Background
- Semantic Hilbert Space—NAACL best explainable NLP paper
  - **Why quantum?**
  - How does it work
  - Any benefits?
- Possible Future works

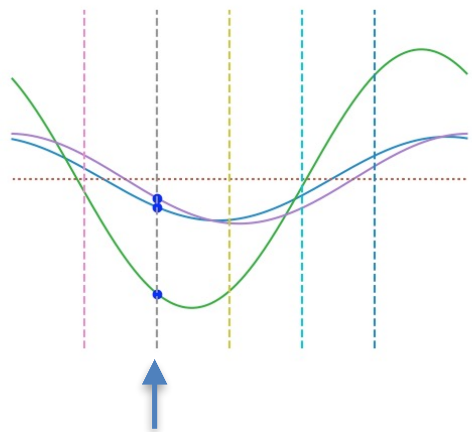
Beyond particles:
understanding words as waves

Word functions for ‘I'



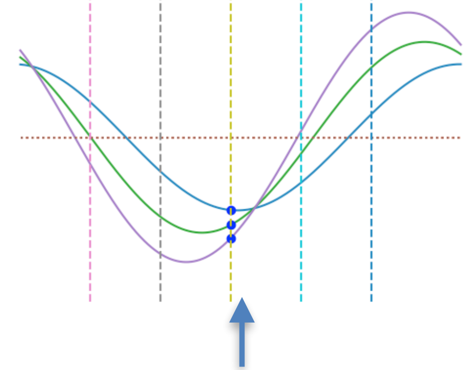
‘I’ is in the 1st position

Word functions for ‘love’



‘love’ is 2nd

Word functions for ‘Amsterdam’



‘Amsterdam’ is 3th

For the sentence ‘I love Amsterdam’

Thanks!

wang@dei.unipd.it