H2020-MSCA-ITN Grant Agreement N. 721321

QUARTZ
Quantum Information Access and Retrieval Theory

Dynamic content monitoring and exploration using vector space (ESR-2)

Benyou Wang

University of Padova

University of Padova, 12/02/2020

ESR2

Dynamic content monitoring and exploration using **vector space**

Vector Space

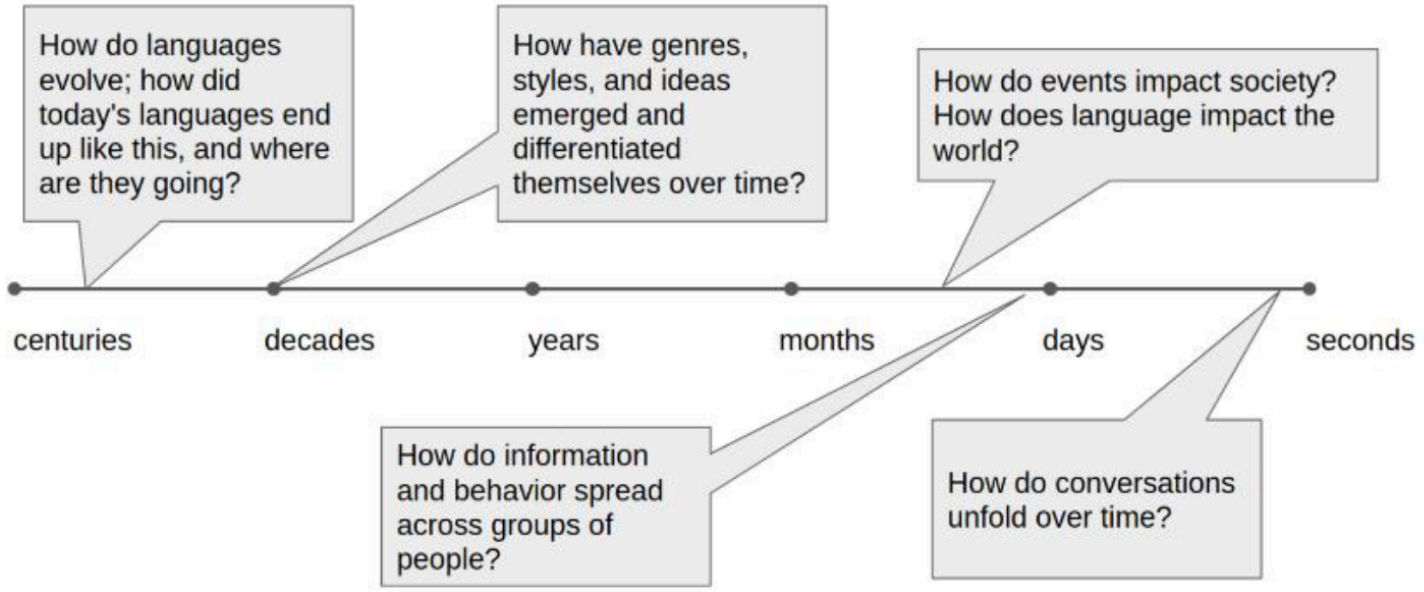
e.g. word embedding (especially complex-valued [1,2]);
semantic Hilbert space (with Quantum probability theory) [3];
tensor space.

[1] **Wang B**., Zhao D., Lioma C., Li Q., Zhang P. and SimonsenJ.G., 2019. Encoding word order in complex embeddings. accepted as ICLR 2020 Spotlight paper
[2] Li Q., **Wang B.** And Melucci M. CNM: An Interpretable Complex-valued Network for Matching. NAACL 2019 Best explainable paper.
[3] **Wang B.**, Li Q., Melucci M. And Song D. Semantic Hilbert Space for Text Representation Learning. WWW 2019.

QUARTZ
Quantum Information Access and Retrieval Theory

Dynamics in Natural language

- Scenarios with extra time/sequential dimension in Natural language



Eisenstein, Jacob. "Measuring and Modeling Language Change." *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Tutorials*. 2019.

Research Plan

- **Vector Space Representation** for **static** text/document
  - Quantum-inspired representation for static text [2,3]

- Extend it to **dynamic** content
  - Modelling sequential order in vector space with wave-like representations [1]
  - Implement it in dynamic corpora, e.g. newspaper, blogs, paper collections, and more generally dynamic content e.g. sequential language modelling, recommendation, soc event detection and dialogue system
  - Evidencing the overall benefit of the proposed methods.

To be done

[1] **Wang B**., Zhao D., Lioma C., Li Q., Zhang P. and SimonsenJ.G., 2019. Encoding word order in complex embeddings. accepted as ICLR 2020 Spotlight paper
[2] Li Q., **Wang B.** And Melucci M. CNM: An Interpretable Complex-valued Network for Matching. NAACL 2019 Best explainable paper.
[3] **Wang B.**, Li Q., Melucci M. And Song D. Semantic Hilbert Space for Text Representation Learning. WWW 2019.

QUARTZ
Quantum Information Access and Retrieval Theory

Current research interest

- Benefit from complex-valued representation
  - Normally, words can be processed linearly
    - $\overrightarrow{\text{men}} - \overrightarrow{\text{women}} + \overrightarrow{\text{queen}} = \overrightarrow{\text{king}}$
  - In some cases, words are not linearly-composable
    - $\overrightarrow{\text{Irovy}} - \overrightarrow{\text{tower}} \neq \overrightarrow{\text{ivory tower}}$

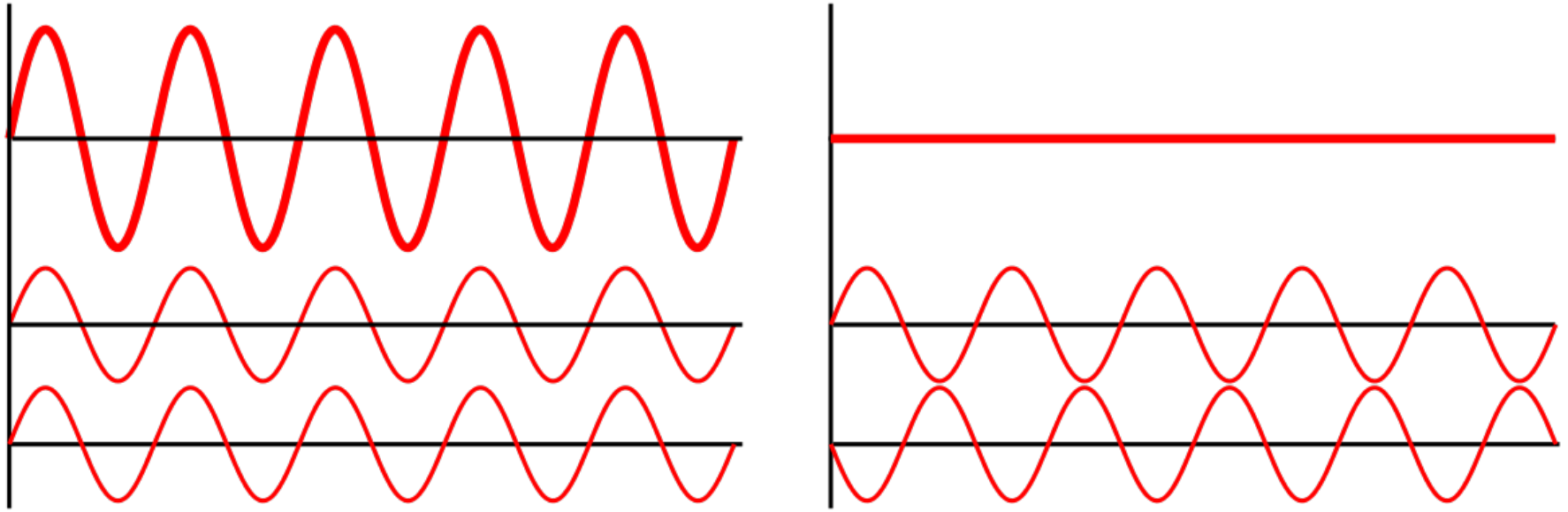
Current research interest

- Benefit from complex-valued representation
  - Normally, words can be processed linearly
    - $\overrightarrow{\text{men}} - \overrightarrow{\text{women}} + \overrightarrow{\text{queen}} = \overrightarrow{\text{king}}$
  - In some cases, words are not linearly-composable
    - $\overrightarrow{\text{Irovy}} - \overrightarrow{\text{tower}} \neq \overrightarrow{\text{ivory tower}}$

Waves with similar phases?

Waves with different phases?

Examples: wave interference



constructive : linear addition

destructive

Words as waves

- Each words as waves with phases
  - Hypothesis: the presentation of a word also depends on its neighbouring words

For words:

$$|\overrightarrow{w_1}|_2 = |\alpha_1 + \beta_1 i|_2 = \sqrt{\alpha_1^2 + \beta_1^2}$$

$$= |r_1 e^{\theta i}|_2 = r$$

For Bigrams

$$|\overrightarrow{w_1} + \overrightarrow{w_2}|_2 = |(\alpha_1 + \alpha_2) + (\beta_1 + \beta_2)i|_2 = \sqrt{(\alpha_1 + \alpha_2)^2 + (\beta_1 + \beta_2)^2}$$

$$= |\sqrt{|r_1|^2 + |r_2|^2 + 2r_1 r_2 cos(\phi_1 - \phi_2)} \times e^{i \arctan\left(\frac{r_1 \sin(\phi_1) + r_2 \sin(\phi_1)}{r_1 \cos(\phi_1) + r_2 \cos(\phi_2)}\right)}|_2$$

$$= \sqrt{|r_1|^2 + |r_2|^2 + 2r_1 r_2 cos(\phi_1 - \phi_2)}$$

Note that there are no extra parameters to represent phrases like bigrams

Potential

Academia

- ✓ Modelling dynamics with well-defined vector space
- ✓ Understanding SOTA models with novel and mathematically-sound angels
- ✓ Investigating Quantum formalisation in representing natural language

industria

- ✓ Better-performed models in textual representation, time-series prediction, event/topic monitoring/conversation system inspired by Quantum theory
- ✓ Compressing models by means of tensor decompositions &tensor networks
- ✓ Investigating wave-based learning algorithms that can be potentially deployed in wave-based computing hardwares like photons (faster and energy-cheap) [1,2]

[1] Lin, Xing, et al. "All-optical machine learning using diffractive deep neural networks." Science 361.6406 (2018): 1004-1008.
[2] Hughes, Tyler W., et al. "Wave physics as an analog recurrent neural network." Science Advances 5.12 (2019):.

QUARTZ
Quantum Information Access and Retrieval Theory

Thanks

Application: Two steps to model language

1 Detect the statistically-noncomposable phrases by

$$\text{score}(w_i, w_j) = \frac{\text{count}(w_i w_j) - \delta}{\text{count}(w_i) \times \text{count}(w_j)}.$$

2 jointly training words and phrase representation