Dynamic content monitoring and exploration using vector spaces

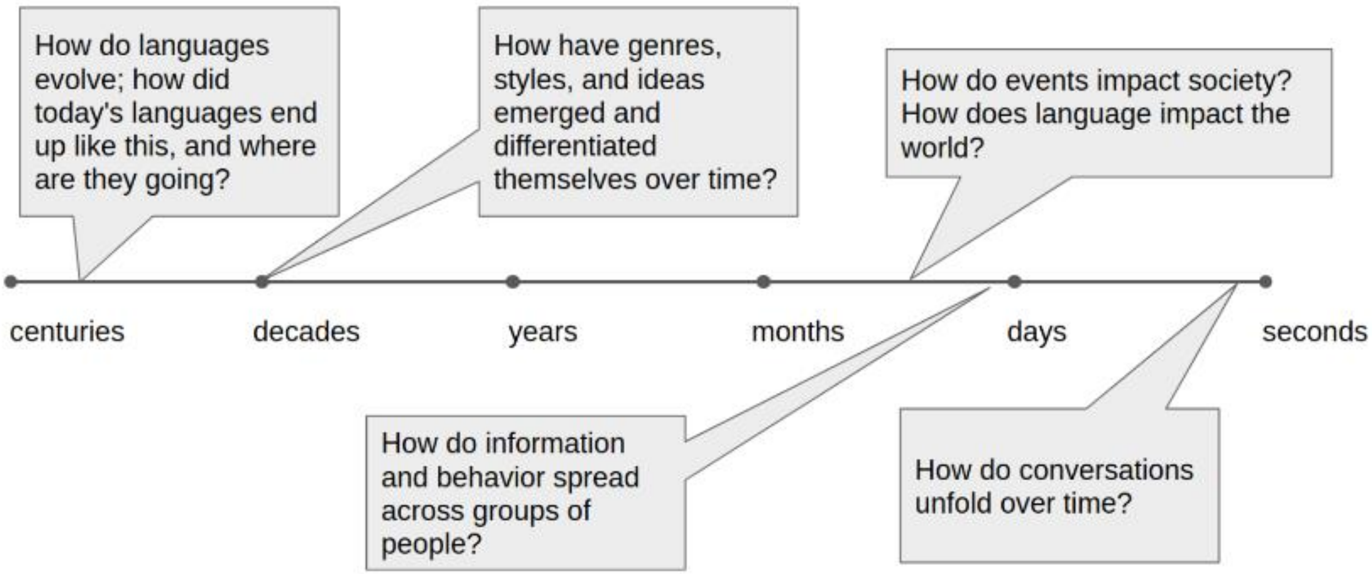
Benyou Wang
University of Padua
Supervised by Massimo Melucci and Emanuele Di Buccio
SIGIR DC, Paris, France, 07/2019

Dynamics

- Sequential data
- Language model and Language generation
- Historical corpora (Google book, ArXiv paper collections)
- Conversation/dialogue
- Recommendation with historical interactions
- Video tracking
- Electrical Healthy records

Focusing on **textual** problem

Taxonomy in textual applications



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.

Why vector spaces?

- ✓ Well-defined properties of vector space and it is naturally used in IR (VSM in IR).
- ✓ Representing words in vector space is a commonly-used paradigm in textual problems [1]

Currently, there is some shortage in modelling dynamic aspects in vector space

[1] Mikolov, Tomas, et al. "Efficient estimation of word representations in vector space." *arXiv preprint arXiv:1301.3781* (2013).

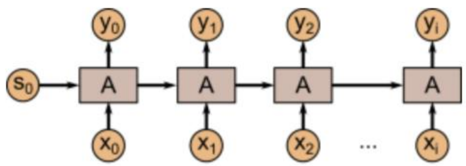
Methods

- ❑ **Inner dynamics**
- ❑ In-between dynamics
- ❑ high-dimensional dynamics

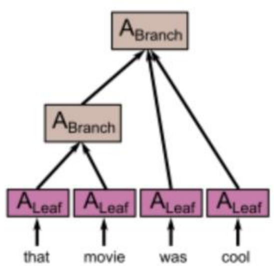
- ✓ **Sequence-sensitive vector space**
- ✓ Evolved vector space
- ✓ Extended high-dimensional vector space

Inner dynamics - Encoding order for a single vector

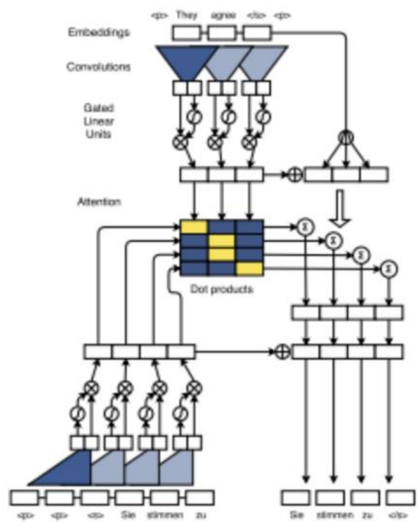
- Problems: position-sensitive vector **without** recurrent architectures



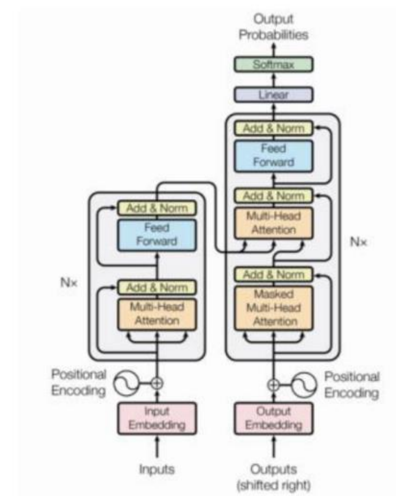
Recurrent Neural Net



Recursive Neural Net



Conv seq2seq



Transformer

Inner dynamics - Encoding order for a single vector

Previous methods : Position Embedding (PE) + Word Embedding (WE)
$\boldsymbol{f}: (\boldsymbol{N}, \boldsymbol{N}) \rightarrow \boldsymbol{R^k}$, e.g., $PE(pos_i) + WE(w_j)$ or $[PE(pos_i); WE(w_j)]$

Extending embedding from a **vector** to a **continuous function** over variable the position (pos)

Technically, $f: (N, N) \rightarrow R^k$ To $\boldsymbol{f}: (\boldsymbol{N}) \rightarrow \boldsymbol{G}\{\boldsymbol{g}; \boldsymbol{g}: \boldsymbol{N} \rightarrow \boldsymbol{R^k}\}$

Word index

position index

Now the question becomes *how to decide the function*

Properties for f

Now, for a specific word w, we have to get its embedding over all the positions, namely a function $\boldsymbol{g_w}: \boldsymbol{N} \rightarrow \boldsymbol{R^k}$

Property 1: Invariant relative-distance transformation
With the any linear transformation $f_{ij}, s.t. f_{ij}(g_w(pos_i)) = g_w(pos_j)$, one should know whether pos_i is in front of pos_j, as well as how far is between pos_i and pos_j, no matter how big $\boldsymbol{pos_i}$ and $\boldsymbol{pos_j}$ is.

Property 2: Boundedness
The function $\boldsymbol{g_w}$ should be bounded, in order to model long-enough sentence

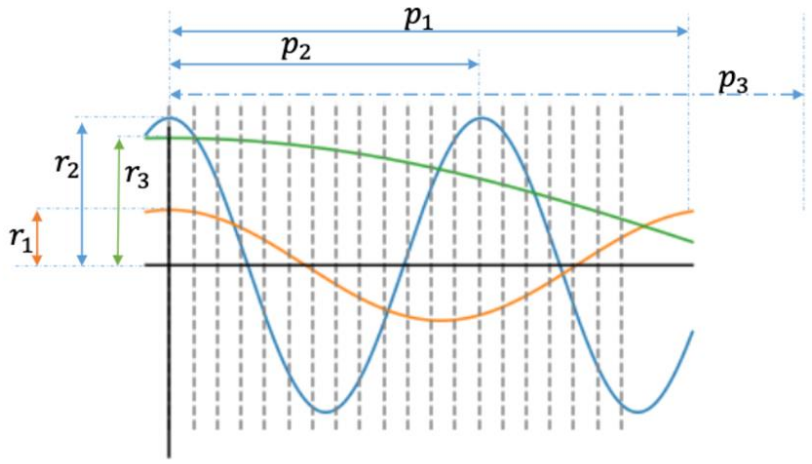
Encoding order in complex embedding

The **only** solution of $\boldsymbol{g_w}: \boldsymbol{N} \to \boldsymbol{R^k}$ to meet the previous properties is

$g_w(pos) = \mathbf{WE_w} \odot \boldsymbol{e}^{\mathbf{i} \cdot \boldsymbol{pos} \cdot \boldsymbol{\lambda}}$

For the k-th dimension, $[g_w(pos)]_k = \mathbf{WE_{w,k}} \cdot \boldsymbol{e}^{\mathbf{i} \cdot \boldsymbol{period_k} \cdot \boldsymbol{pos}}$
With Euler' formula, we can get

$[g_w(pos)]_k = WE_{w,k} \cdot [\mathbf{cos}(\boldsymbol{period_k} \cdot \boldsymbol{pos}) + \boldsymbol{i}\, \mathbf{sin}(\boldsymbol{period_k} \cdot \boldsymbol{pos})]$



Methods

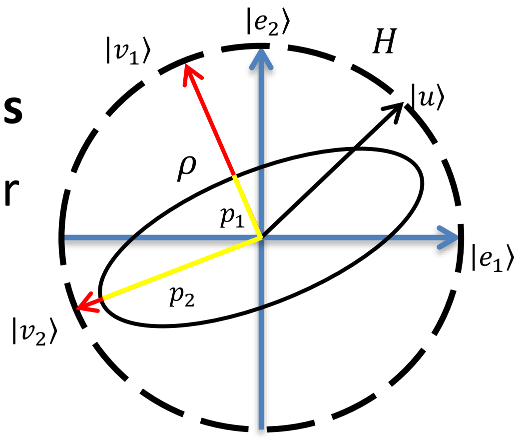
- ❑ Inner dynamics
- ❑ **In-between dynamics**
- ❑ high-dimensional dynamics

- ✓ Sequence-sensitive vector space
- ✓ **Evolved vector space**
- ✓ Extended high-dimensional vector space

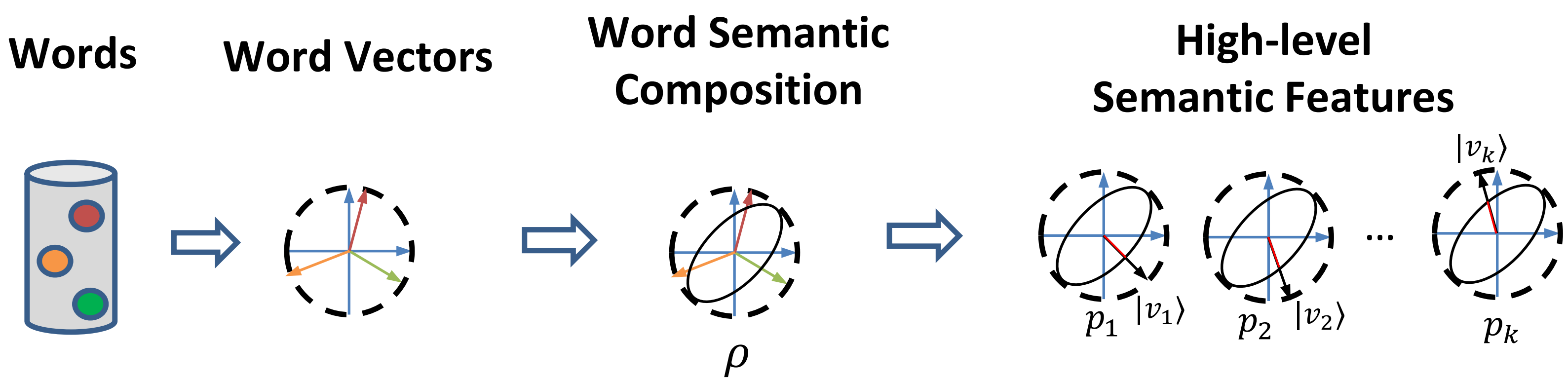
Quantum probability

- Classical probability
  - Set-based probability theory
  - Events are limited to be discrete and mutually-exclusive

- Quantum probability
  - Projective geometry based probability theory
  - Events, which is defined in a complex **continuous** vector space, can be represented as arbitrary vector

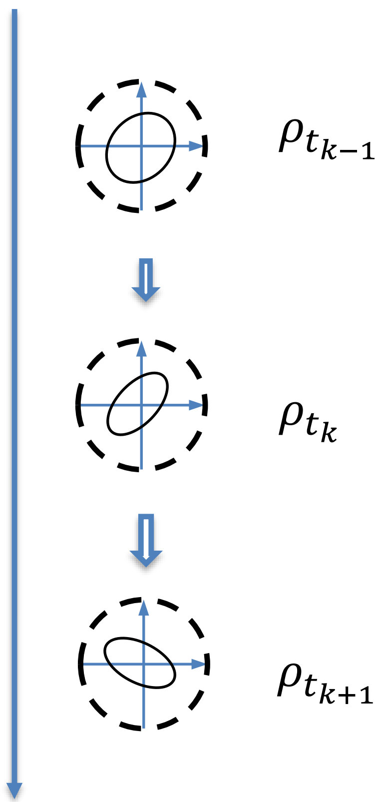


Density matrix in vector space



Li, Qiuchi *, **Benyou Wang ***, and Massimo Melucci. "CNM: An Interpretable Complex-valued Network for Matching." NAACL. 2019. **Best explainable NLP paper**

Evolved density matrix



Historical corpora
Dialogue system

$$\rho_{t_{k+1}} = u\rho_{t_k}u^*, \text{ where u is unitary}$$

$$\frac{d\rho_{t_k}}{dt} = f(\rho, t_k)$$

Advantages of Evolved density matrix

- Better interpretability from quantum probability [1]
- Better optimization based on unitary transformation [2]
- Linking with neural (Schrödinger) differential equation [3]

[1] Li, Qiuchi *, **Benyou Wang ***, and Massimo Melucci. "CNM: An Interpretable Complex-valued Network for Matching." NAACL. 2019. **Best explainable NLP paper**
[2] Arjovsky, Martin, Amar Shah, and Yoshua Bengio. "Unitary evolution recurrent neural networks." *ICML*. 2016.
[3] Chen, Tian Qi, et al. "Neural ordinary differential equations." *NIPS*. 2018.

Future works

- Investigating the dynamic aspects in high-dimensional vector space, i.e., tensor space

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

Cohen, Nadav, Or Sharir, and Amnon Shashua. "On the expressive power of deep learning: A tensor analysis." *COLT*. 2016.
Khrulkov, Valentin, Alexander Novikov, and Ivan Oseledets. "Expressive power of recurrent neural networks." *arXiv preprint arXiv:1711.00811* (2017). ICLR 2018

Publications

1. Qiuchi Li*, **Benyou Wang***, Massimo Melucci. A Complex-valued Network for Matching. **NAACL 2019**, **Best Explainable NLP Paper**
2. **Benyou Wang***, Qiuchi Li*, Massimo Melucci, Dawei Song. Semantic Hilbert Space for Text Representation Learning. **WWW 2019**
3. Wei Zhao*, **Benyou Wang***, Min Yang, Jianbo Ye, Zhou Zhao, Xiaojun Chen, Ying Shen.. Leveraging Long and Short-term Information in Content-aware Movie Recommendation via Adversarial Training. **IEEE Transactions on Cybernetics** (TOC), 2019
4. Peng Zhang, Zhan Su, Lipeng Zhang, **Benyou Wang** , Dawei Song. 2018. A Quantum Many-body Wave Function Inspired Language Modeling Approach. **CIKM 2018**
5. Wei Zhao, **Wang Benyou** , Jianbo Ye, Yongqiang Gao, Min Yang, Xiaojun Chen, PLASTIC: Prioritize Long and Short-term Information in Top-n Recommendation using Adversarial Training, **IJCAI 2018**
6. Wei Zhao, **Wang Benyou** , Jianbo Ye, Min Yang, Zhou Zhao, Ruotian Luo, Yu Qiao A Multi-task Learning Approach for Image Captioning, **IJCAI 2018**
7. Zhang Peng, Niu Jiabing, Su Zhan, **Wang Benyou** et al. End-to-End Quantum-like Language Models with Application to Question Answering **AAAI 2018**
8. Wang Jun, Yu Lantao, Zhang Weinan, Gong Yu, Xu Yinghui, **Wang Benyou** , Zhang Peng, Zhang Dell. IRGAN: A Minimax Game for Unifying Generative and Discriminative Information Retrieval Models. **SIGIR 2017**. **Best Paper Award Honourable Mentions**.
9. Huang Xin, Wei zhao, Wang Benyou, Rui Zhao. Recommendation System and Deep Learning, Tsinghua University Press, in Chinese.

Thanks