Representations and their Matching
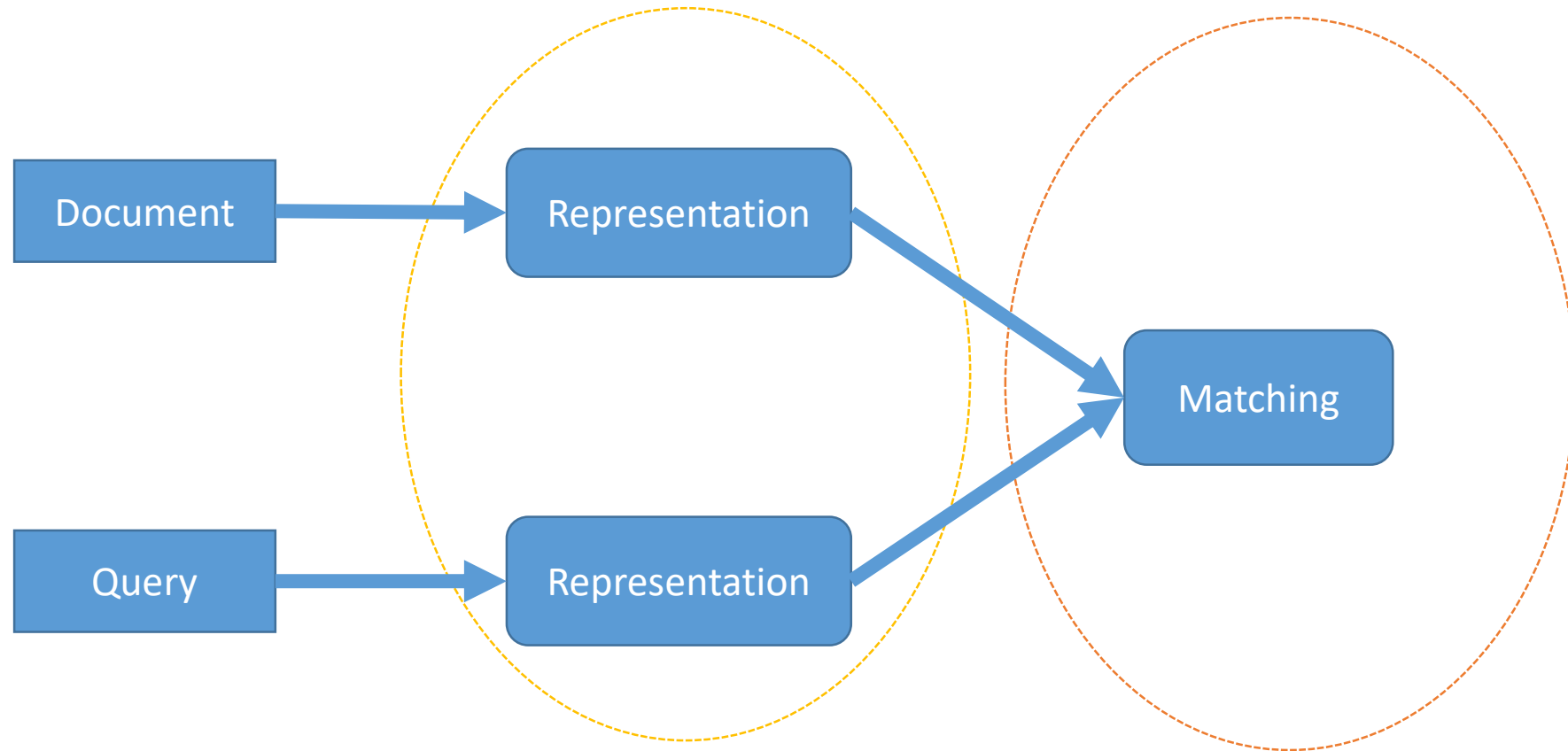
Overview of my research

Benyou Wang

Supervised by Prof. Massimo Melucci

Examples for Representation & Matching

Refer to Benyou wang's homepage, https://wabyking.github.io/talks/textzoo.pdf

Key concerns

- ## How to build a good **representation**
  - Language model   [End-2-End QLM AAAI 2018] [QMWF-LM AAAI 2018]
  - Neural representation   [Complex Word Embedding ACL 2018][TextZoo Arxiv 2018][Quantum Attention]
  - Multi-modality   [Image caption IJCAI2018][Sentimental analysis TCS]
- ## How to make a good matching
  - Ad hoc retrieval   [Quantum Query Expansion Entropy 2018]
  - Question Answering   [End-2-End QLM AAAI 2018]
  - GAN for Matching   [IRGAN SIGIR 2017]
  - Recommendation   [Long + short-term Profile IJCAI 2018]
  - Customer Service   [QA NLPCC 2016&NLPCC 2018]

  - ### Tools
    - Quantum Concepts or Quantum-inspired method
    - GAN/Multitask/DNN/Eye tracking

Faster language model with CNN

Wh
- -at
- -en
- -ich
- -ere

Model	Test PPL	Hardware
Sigmoid-RNN-2048 (Ji et al., 2015)	68.3	1 CPU
Interpolated KN 5-Gram (Chelba et al., 2013)	67.6	100 CPUs
Sparse Non-Negative Matrix LM (Shazeer et al., 2014)	52.9	-
RNN-1024 + MaxEnt 9 Gram Features (Chelba et al., 2013)	51.3	24 GPUs
LSTM-2048-512 (Jozefowicz et al., 2016)	43.7	32 GPUs
2-layer LSTM-8192-1024 (Jozefowicz et al., 2016)	30.6	32 GPUs
BIG GLSTM-G4 (Kuchaiev & Ginsburg, 2017)	23.3*	8 GPUs
LSTM-2048 (Grave et al., 2016a)	43.9	1 GPU
2-layer LSTM-2048 (Grave et al., 2016a)	39.8	1 GPU
GCNN-13	38.1	1 GPU
GCNN-14 Bottleneck	31.9	8 GPUs

Yin W, Kann K, Yu M, et al. Comparative study of cnn and rnn for natural language processing[J]. arXiv preprint arXiv:1702.01923, 2017.
https://github.com/wabyking/Gated_CNN_for_language_Modeling,
Dauphin Y N, Fan A, Auli M, et al. Language Modeling with Gated Convolutional Networks**. ICML 2017**: 933-941.

TextZOO [Benyou et.al Arkiv 2018]

Models

- ✓ FasText
- ✓ CNN (Kim CNN, Multi-Layer CNN, Multi-perspective CNN, Inception CNN)
- ✓ LSTM (BILSTM, StackLSTM, LSTM with Attention)
- ✓ Hybrids between CNN and RNN (RCNN, C-LSTM)
- ✓ Attention (Self Attention / Quantum Attention)
- ✓ Transformer - Attention is all you need
- ✓ Capsule
- ✓ Quantum-inspired NN
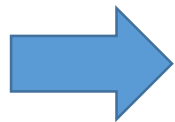- ➤ ConS2S
- ➤ Memory Network

Datasets

- **IMDB**
- **MR**
- CR
- MPQA
- SST1
- SST2
- Subj
- TREC

Wang B, Wang L, Wei Q. TextZoo, a New Benchmark for Reconsidering Text Classification[J]. arXiv preprint arXiv:1802.03656, 2018.
https://github.com/wabyking/TextClassificationBenchmark

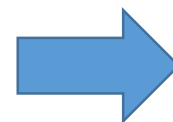# Understand the content of a movie [Wei et.al IJCAI 2018]

QQ空间

Keyframe extraction
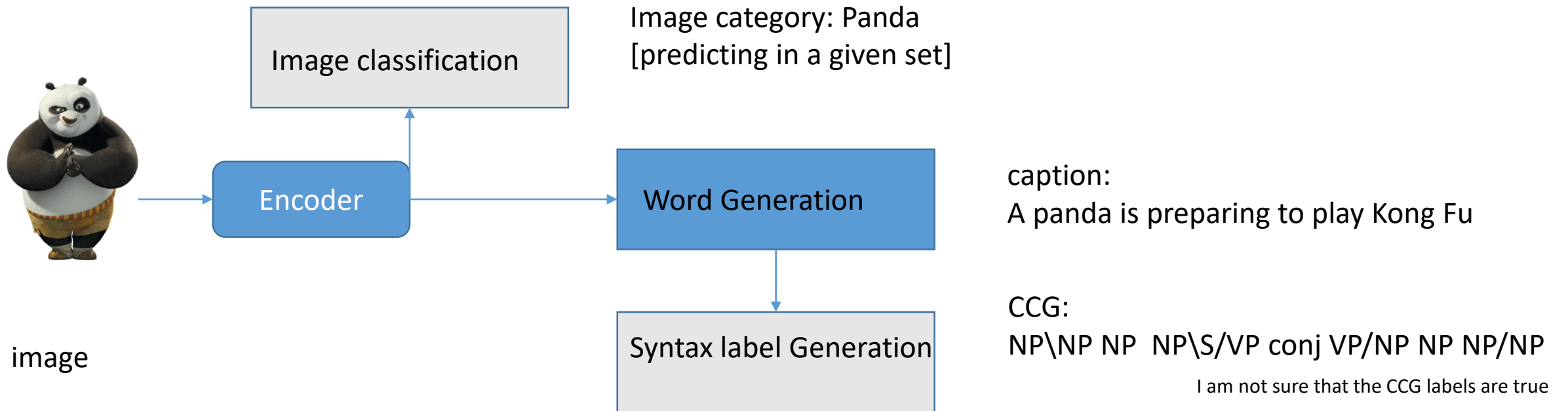
Key frames

Image Caption

Textual features

A panda is preparing to play Kong Fu

A panda is eating noodles on the table

Two pandas are laughing

Multi-task for encoder and decoder

Image classification

Image category: Panda
[predicting in a given set]

Encoder

Word Generation

caption:
A panda is preparing to play Kong Fu

CCG:
NP\NP NP NP\S/VP conj VP/NP NP NP/NP

Syntax label Generation

I am not sure that the CCG labels are true

image

Wei Zhao, Benyou Wang, et.al. A Multi-task Learning Approach for Image Captioning ,IJCAI 2018

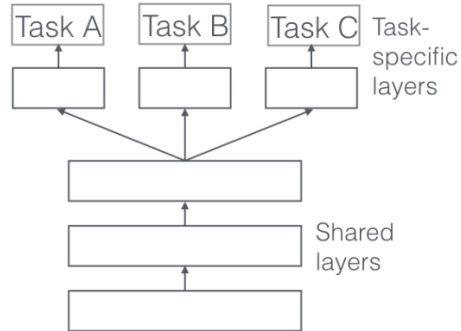# Multi-task approach

Example: Using **query logs** to infer the user profile like age, sex and education background

Google

Quantum Theory
Information Retrieval
How to cook a nice Chicken
Travel in Europe
How to enjoy a academic career with the professor

Age: 0-10
       10-20
       20-40

sex: female
     man

education: Doctor
           Master
           Bachelor
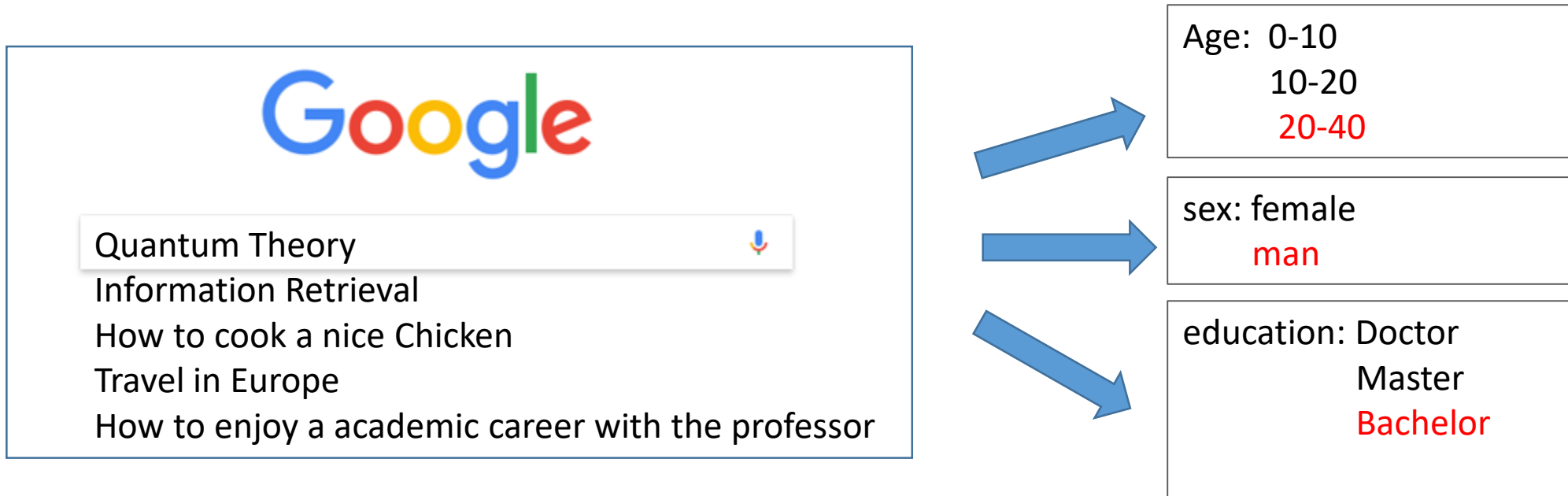
Two Typical Multi-task Paradigms

Hard parameter sharing
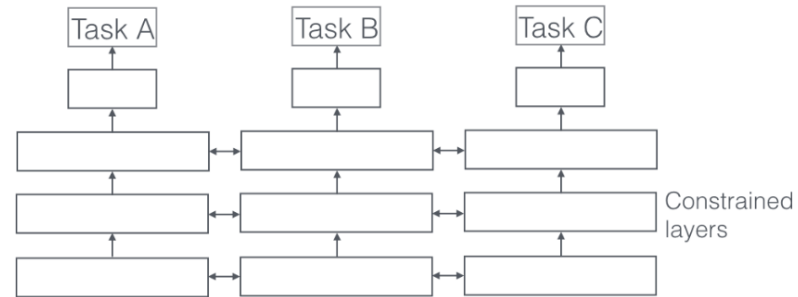
Soft parameter sharing

The distance between the parameters of the model is then regularized in order to encourage the parameters to be similar

http://ruder.io/multi-task/
Caruana, R. "Multitask learning: A knowledge-based source of inductive bias. **ICML**. 1993.
Duong, L., Cohn, T., Bird, S., & Cook, P. Low Resource Dependency Parsing: Cross-lingual Parameter Sharing in a Neural Network Parser. **ACL-NAACL 2015**

Representation inspired by Quantum

- Quantum Probability Space

- Quantum Many-body wave function and Tensor language model

- Quantum Capsule Models (Using direction, instead of numerical number)

- Quantum two-state Formalism

Melucci M. An Algorithm to Calculate a Quantum Probability Space[J]. arXiv preprint arXiv:1710.10158, 2017.
Zhan Su, Peng Zhang, Lipeng Zhang, Benyou Wang, et.al. A Quantum Many-body Wave Function Inspired Language Modeling Approach, submitted to CIKM 2018.
Pestun V, Vlassopoulos Y. Tensor network language model[J]. arXiv preprint arXiv:1710.10248, 2017.
Sabour S, Frosst N, Hinton G E. Dynamic routing between capsules[C]//Advances in Neural Information Processing Systems. 2017: 3856-3866.

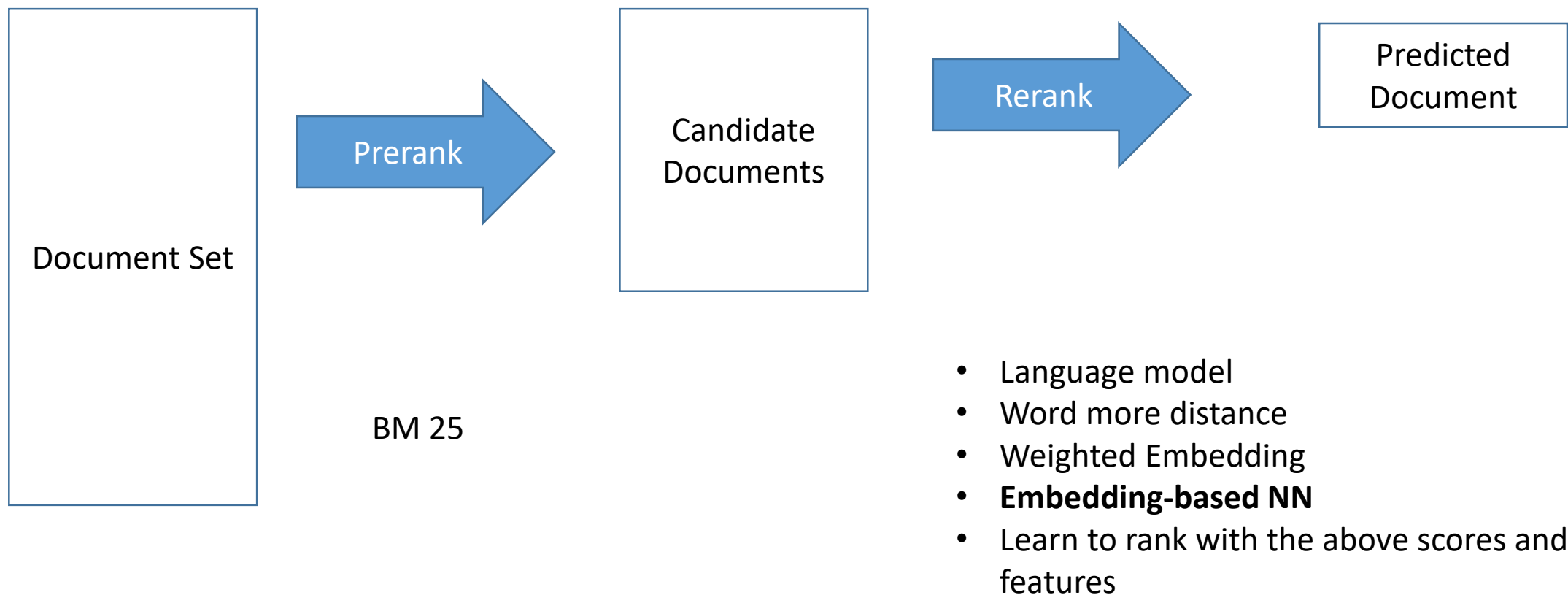# Key concerns

- How to make a good matching
    - Ad hoc retrieval   [Quantum Query Expansion Entropy 2018]
    - Question Answering   [End-2-End QLM AAAI 2018]
    - GAN for Matching   [IRGAN SIGIR 2017]
    - Recommendation   [Long + short-term Profile IJCAI 2018]
    - Customer Service   [QA NLPCC 2016&NLPCC 2018]

Customer Service in Tencent

Given a set of Frequent Question Answer Pairs, and answer a new question from the give QA collection.

Document Set → **Prerank** → Candidate Documents → **Rerank** → Predicted Document

BM 25

- Language model
- Word more distance
- Weighted Embedding
- **Embedding-based NN**
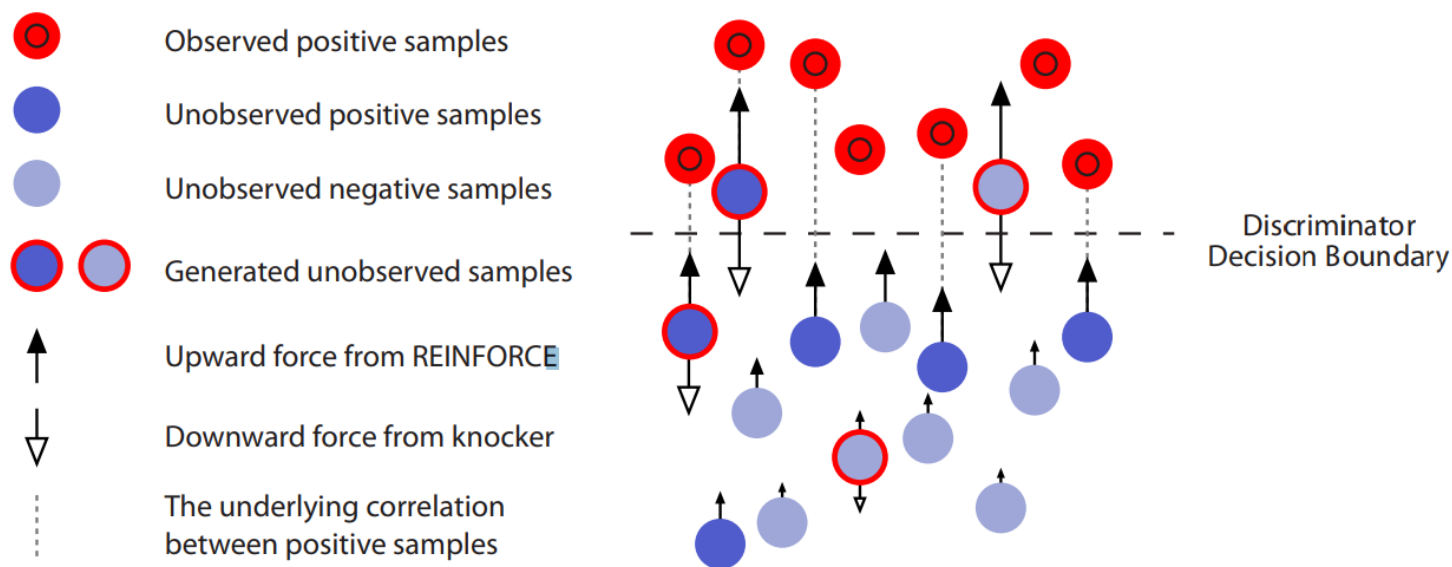- Learn to rank with the above scores and features
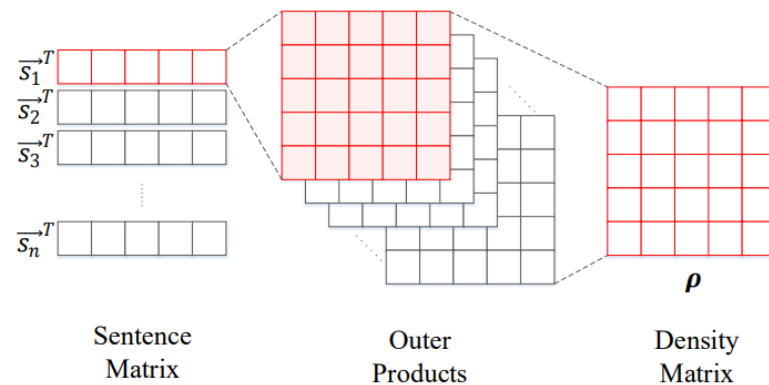
QA matching tasks from a industrial view

- Count-based VS Embedding-based
  - Count-based bag-of-word models are more robust
  - Embedding-based models needs supervised corpus.

- If you have **enough more high-quality supervised matching pair**. It should achieve much better performance

Wang Benyou, Niu Jiabing, Ma Liqun, Zhang Yuhua, Zhang Lipeng, Li Jinfei, Zhang Peng Song, D. . A Chinese Question Answering Approach Integrating Count-Based and Embedding-Based Features. ICCPOL-NLPCC . December, 2016
Su Zhan, **Wang Benyou**, Niu Jiabin, Tao Shuchang, Zhang Peng, Song Dawei. Enhanced Embedding based Attentive Pooling Network for Answer Selection. NLPCC 2017

IRGAN [Jun et.al. SIGIR 2017]

Adjust the original unsupervised models via the feedback from the supervised ones

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** .

End2end Language model [Peng et.al AAAI 2018]

Sentence Matrix   Outer Products   Density Matrix
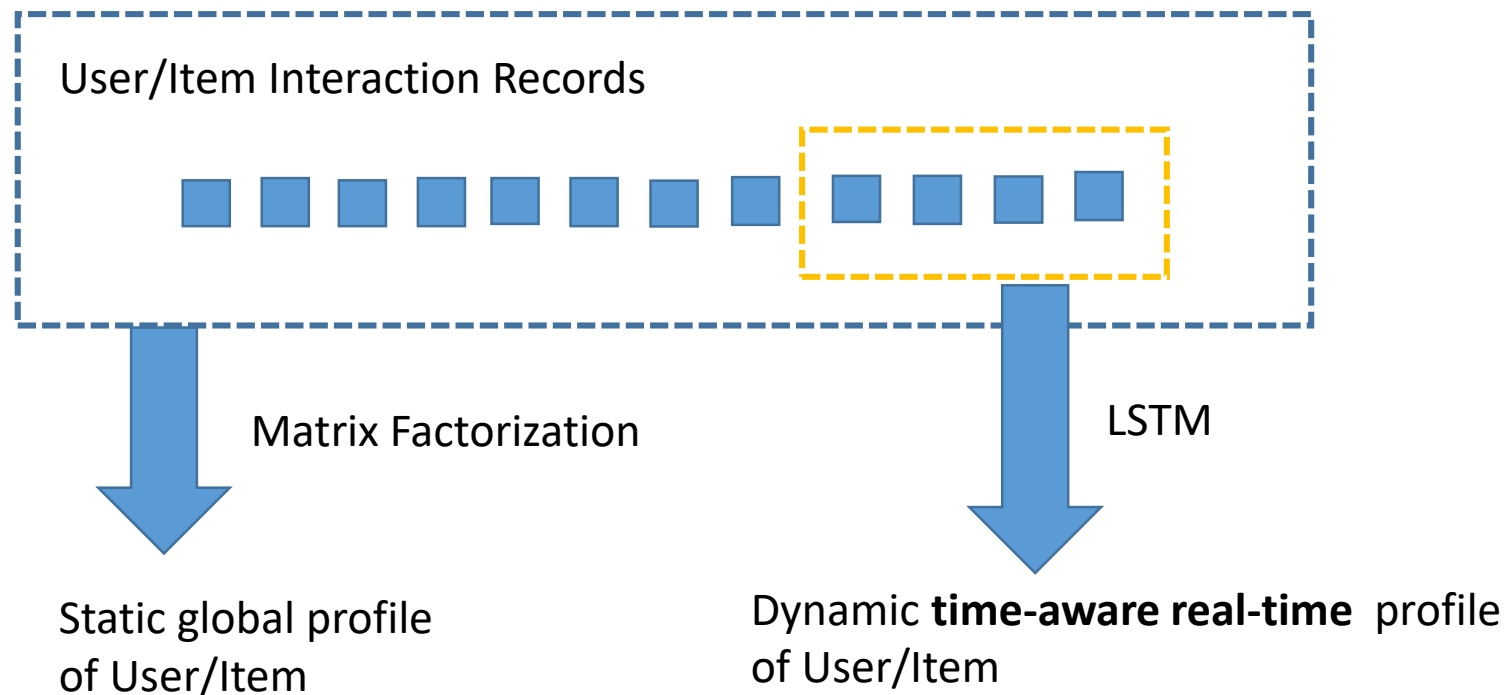
Matching with two matrices

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

*More **non-quantumnic**, the better in performance*

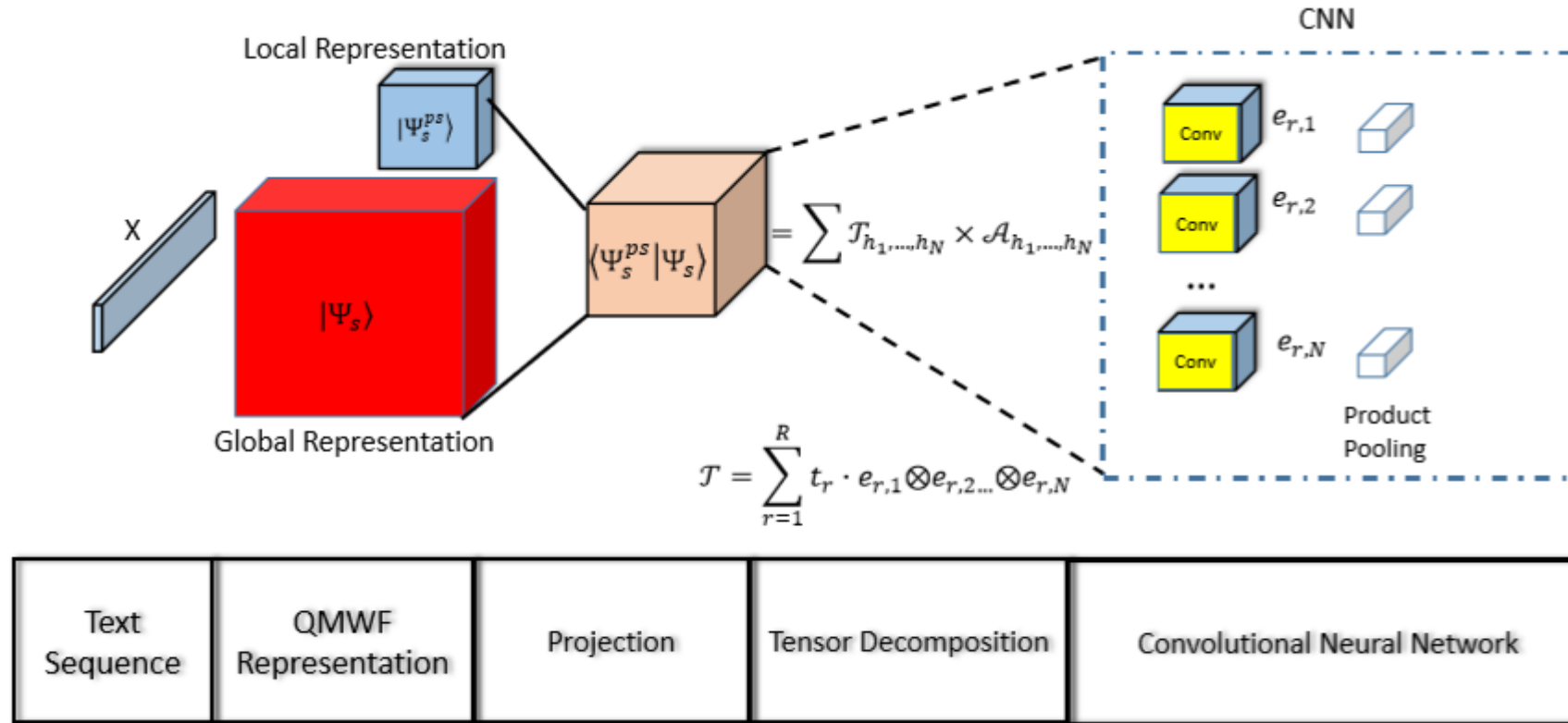1.Zhang Peng, Niu Jiabing, Su Zhan, **Wang Benyou** et al. End-to-End Quantum-like Language Models with Application to Question Answering **AAAI 2018**

Long/short profile for Item/User [Wei et.al IJCAI 2018]

User/Item Interaction Records

Matrix Factorization

LSTM

Static global profile
of User/Item

Dynamic **time-aware real-time** profile
of User/Item

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**

Long/short profile for Item/User [Wei et.al IJCAI 2018]

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**

Quantum Many body language for NN

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

Future

I am Open with the research topics

- Quantum Probability Space
- Contextual Quantum language model in Dynamics
- Capsule Network with Quantum mechanism
- Develop unsupervised IR models with adversarial method
- ….

Thanks