YouroQNet

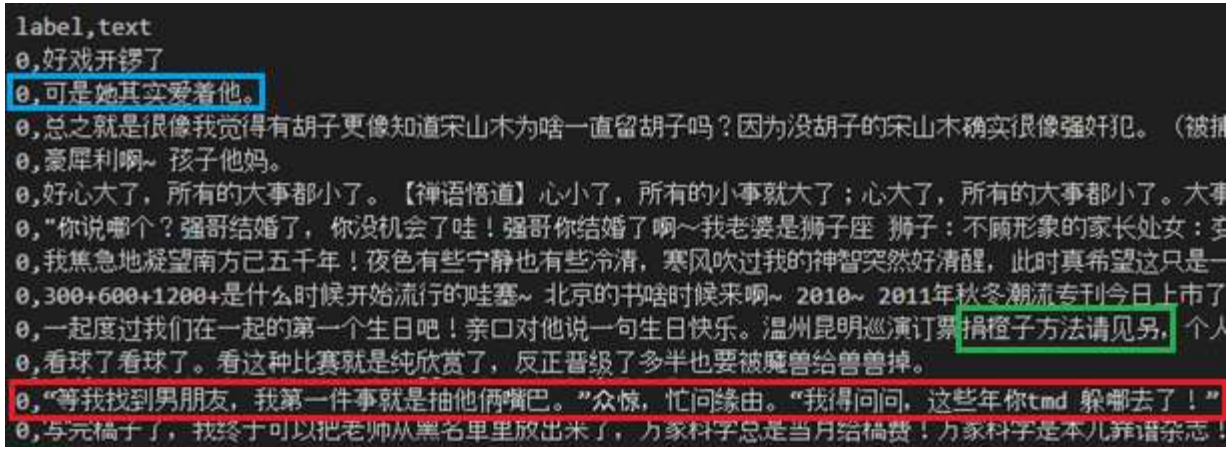
Quantum Text Classification with Context Memory

Team: QwQ

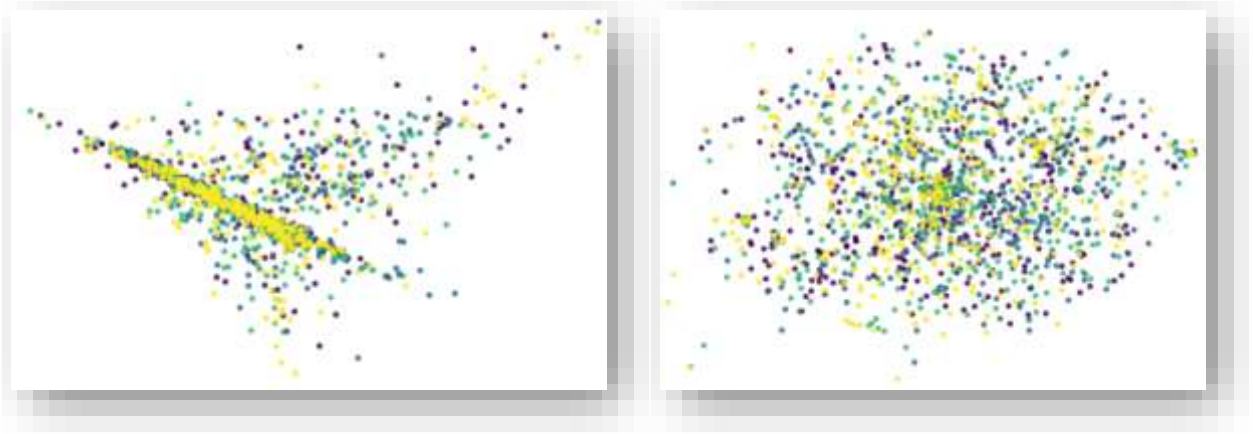
Reporter: Armit

Text Classification in a Quantum Manner

- what the problem is
  - NLP sentiment comprehension
  - the dataset just holy shit
- how current methods work
  - TextCNN, TextRNN, BERT
  - QNLP , QSANN
- nevertheless our contributions
  - heuristic k-gram tokenizer
  - YouroQNet
  - computational analysis



ambiguous example, wrong label and informal pragmatics

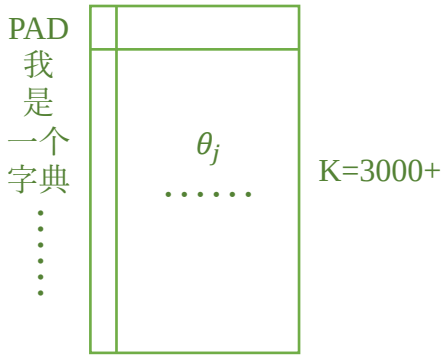


PCA & TSNE over word-level TF-IDF

kgram tokenizer

这/是/一/个/例/子 logp = -50.4
这是/一个/例子 logp = -46.4
这是/一个/例/子 logp = -39.0
……

word embedding



aligner

(split or pad to model length)

这 / 是 / 一个
是 / 一个 / 例子

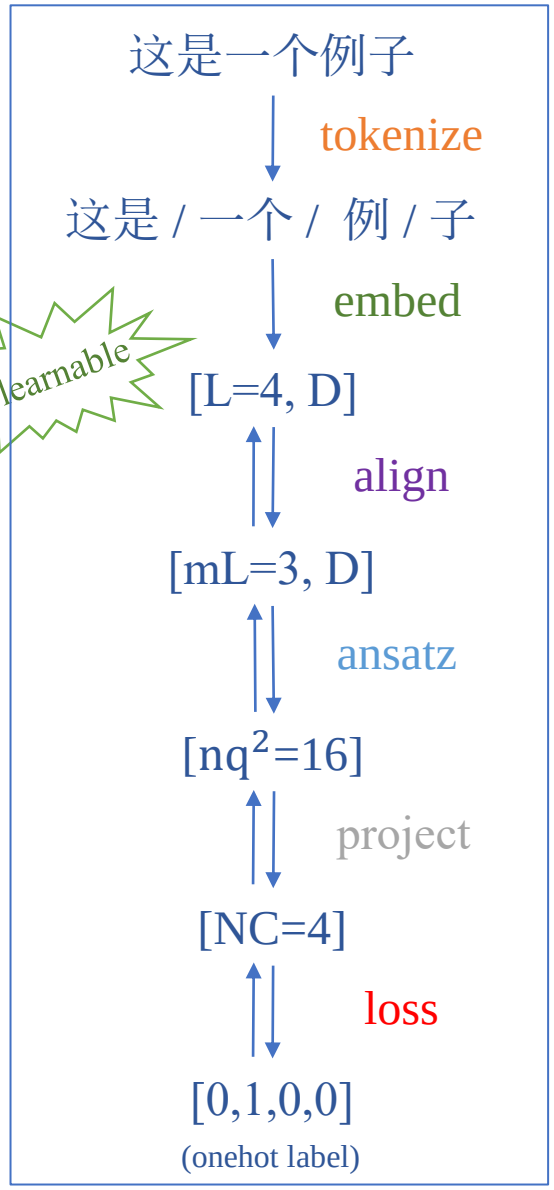
loss

(binary cross entroy)

$$= y\log(x) + (1-y)\log(1-x)$$

$x \in [0,1], y \in \{0,1\}$

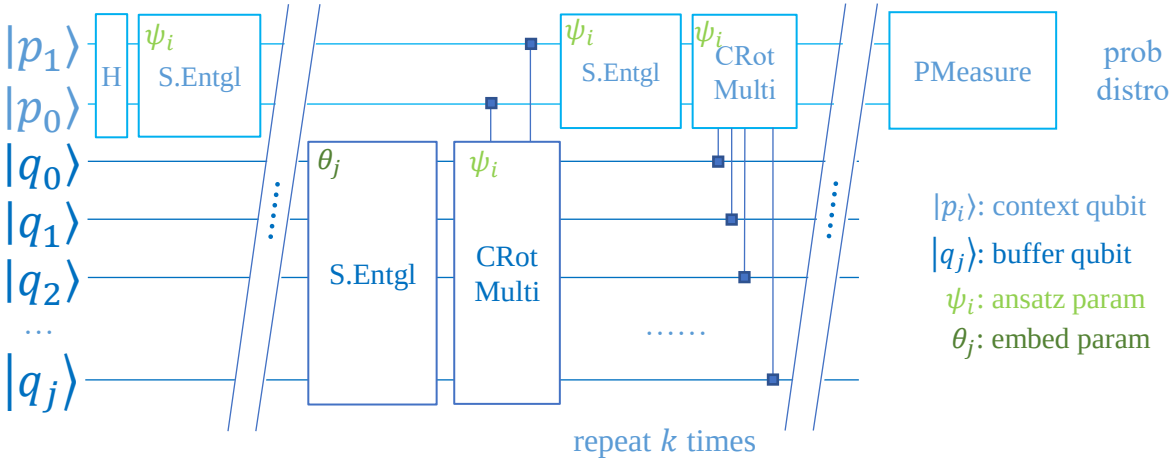
Text Classification



ansatz

(YouroQNet)

n_qubit = model_length + n_class



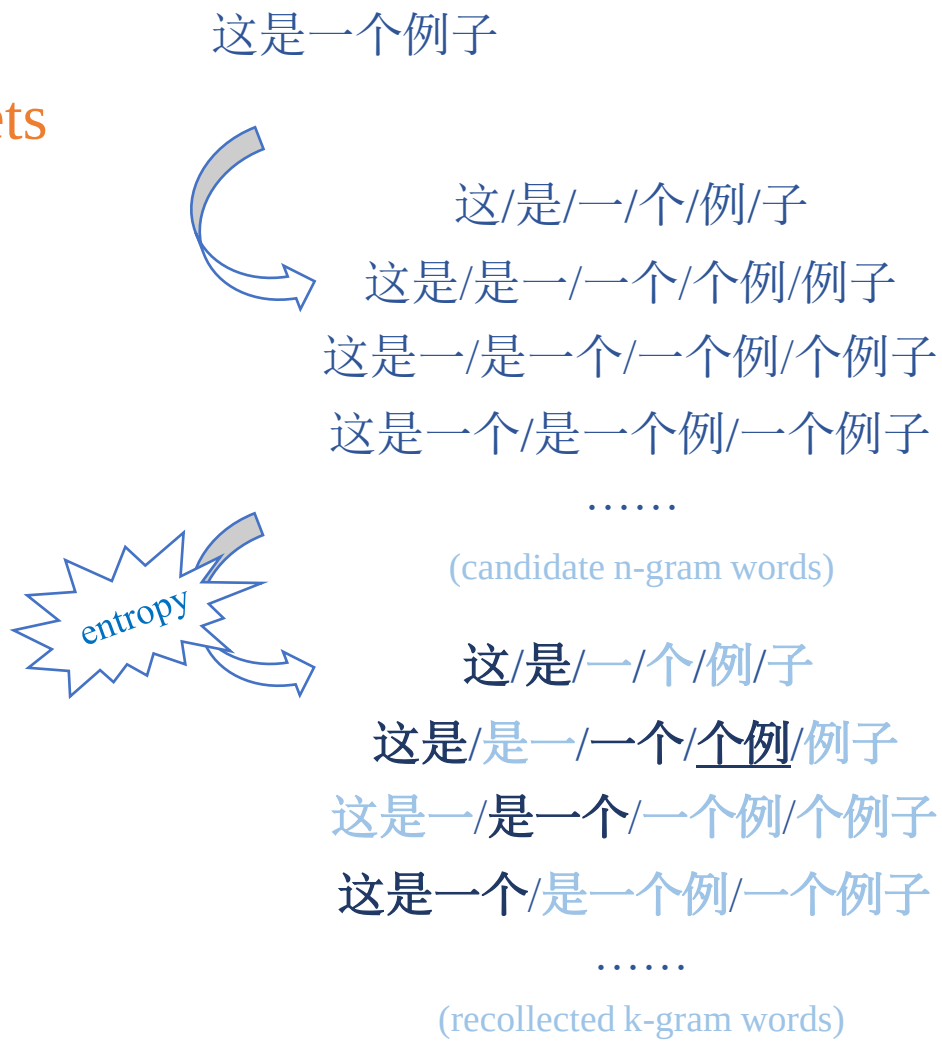
Read the circuit:

- $|p_i\rangle$ is current context, $|q_j\rangle$ is buffer for incoming data
- param ψ_i is for ansatz, θ_j is from word embed
- initialize context with H gate
- bias context via a SEC (Strong Entangle Circuit)
- repeat k times for sequence comprehension
  - load data via a SEC
  - write to context via a CMC (CRot Multi Circuit)
  - digest context via a SEC
  - read from context via a CMC
- prob measure on context qubits

Content Table

Entropy-based Heuristic k-gram Tokenizer

- why tokenizing
  - not necessary, even not work on small datasets
  - reduce input length for limited qubits
- from fixed n-gram to adaptive k-gram
  - build several n-gram vocabs from the corpus
    - word presence probability
  - tokenize corpus by longest match
    - bidirectional beam search
  - recollect all tokenized words as new vocab
    - can be iterative if needed



Tokenizer: concrete example

- processing pipeline
  - text normalization
  - make char and 2/3/4/5-gram vocab
    - filter Chinese words for n-gram
    - truncate size by min_freq=3
    - calculate word presence probability
  - tokenize the corpus with merged vocabs
    - build trie tree for longest match
    - bi-directional beam search with n_beam=3
    - gather all tokenized tokens
    - truncate size by min_freq=5
  - maximize $p(W) = \prod_i p(w_i)$
    - $p(这)p(是)p(一)p(个例)p(子)$
    - $p(这是)p(一个)p(例)p(子)$
    - …

```
s = R_DEADCHAR.sub('', s)
s = R_CJK_PERIOD.sub('.', s)
s = R_CJK_PUASE.sub(',', s)
s = wchar_to_char(s)
s = s.lower()
s = R_WHITESPACE.sub(' ', s)
s = R_NUMBER.sub('0', s)
s = try_concat(s)
s = fold_triple(s)
```

text norm

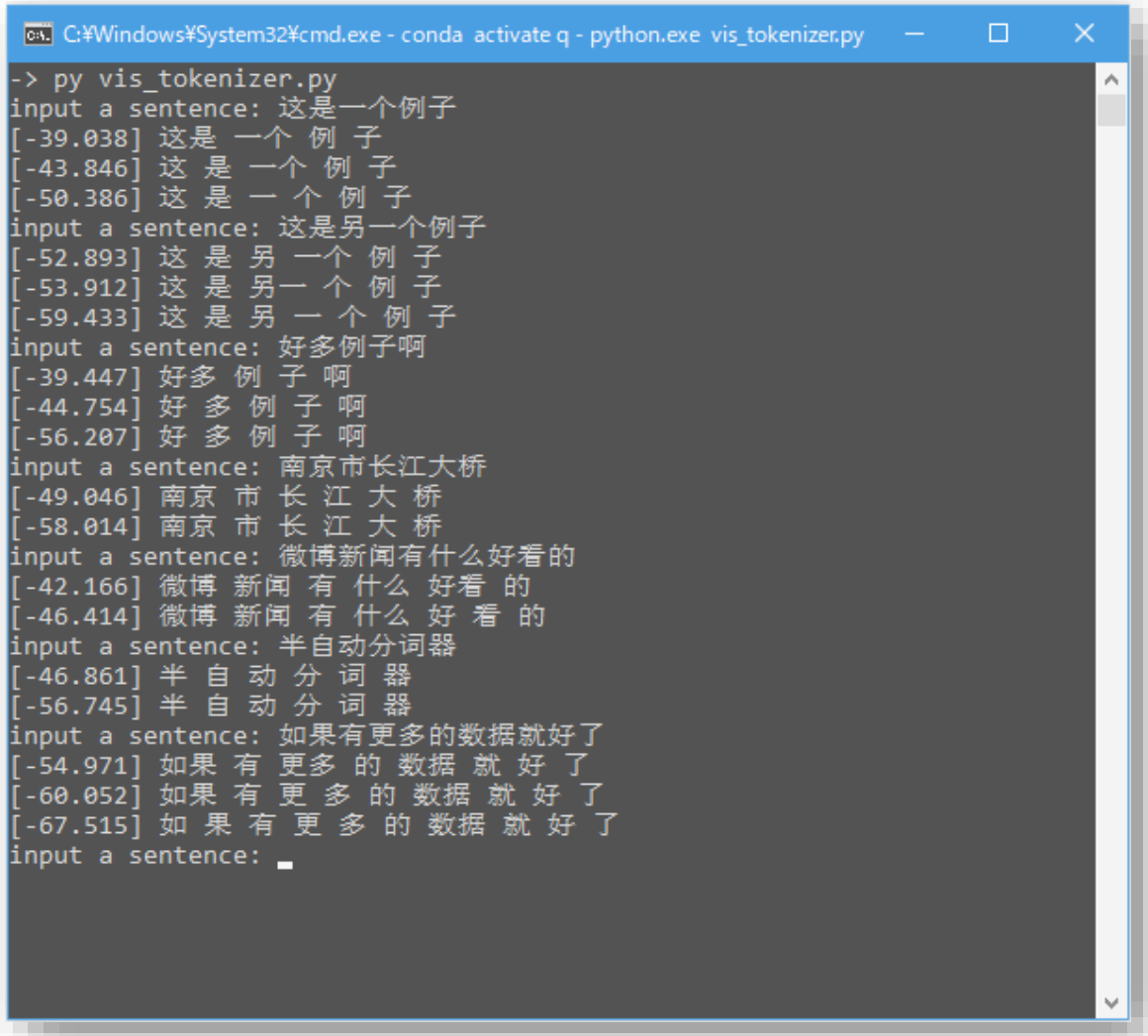
n-gram	n_vocab	n_vocab trunc.
char	3345	-
2-gram	34363	4438
3-gram	48720	1264
4-gram	45478	238
5-gram	38812	44
k-gram	4934	2713
k-gram+	6376	2899

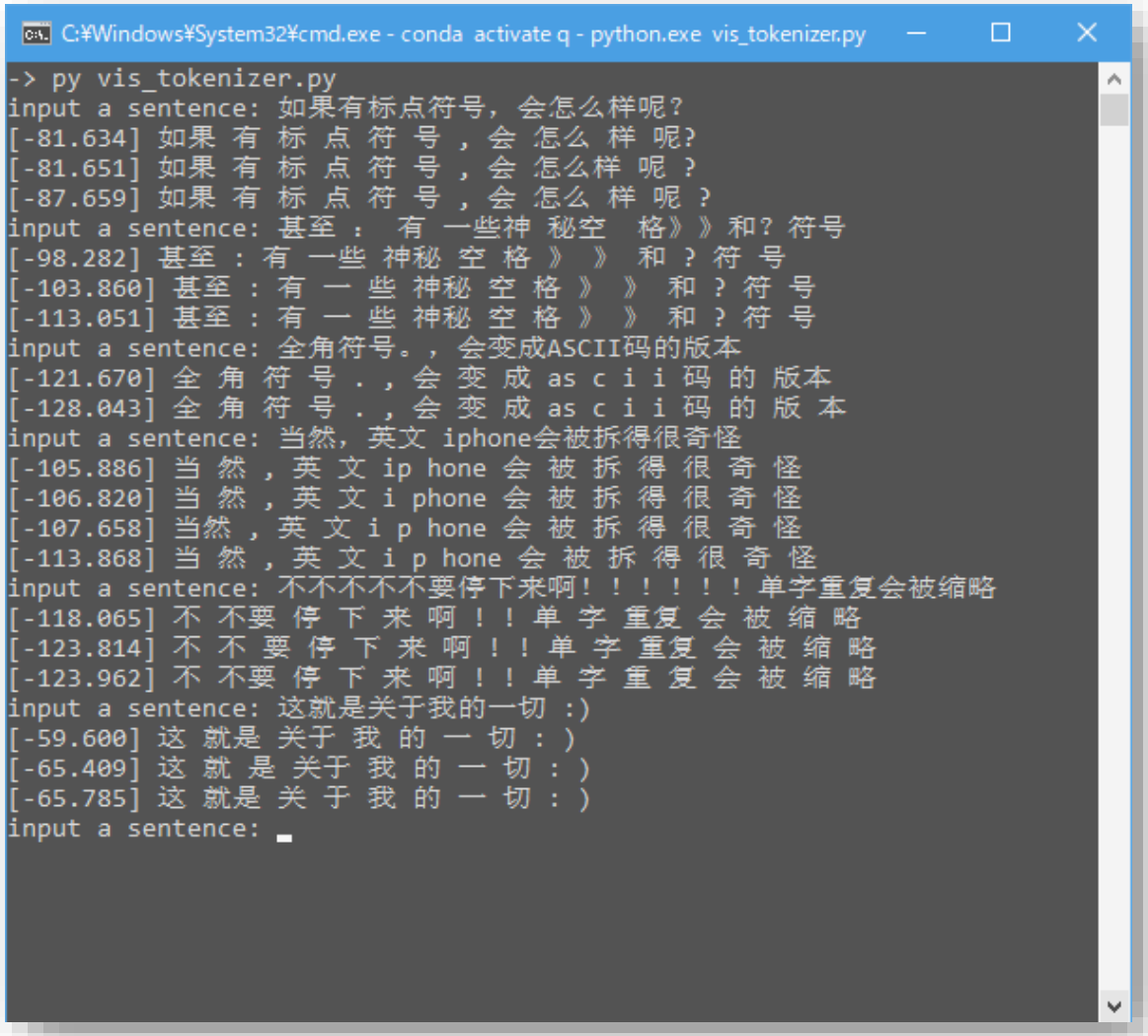
n/k-gram vocabs size

1	,	3822	一个	201	一个人	58	心不开心	32	开心不开心	32
2	.	2895	我们	176	的时候	55	开心不开	32	不开心不开	31
3	的	2717	什么	158	不知道	39	不开心不	31	心不开心不	30
4	我	1587	不是	125	不开心	37	分享图片	17	黑板上画了	5
5	!	1566	我的	121	为什么	33	我不知道	10	在黑板上画	5
6	了	1437	就是	119	心不开	32	内牛满面	10	不知道什么	5
7	是	1290	这个	118	开心不	32	什么时候	10	知道什么叫	4
8	不	1212	自己	115	自己的	31	生日快乐	9	爱情就像便	4
9	一	1136	哈哈	105	这样的	27	黑板上画	7	有机会获得	4
10	0	983	今天	103	是一个	24	让我们再	7	情就像便便	4
11	人	789	的人	102	是不是	23	新浪微博	7	发光的友情	4
12	~	783	大家	100	分享图	22	不是因为	7	这样的男人	3
13	你	735	真的	96	演唱会	21	这就是我	6	这就是传说	3
14	有	730	回复	91	一定要	21	轉發微博	6	还在黑板上	3
15	这	681	没有	89	真的很	19	知道什么	6	观世音菩萨	3

n-gram vocabs

Tokenizer: interactive demo



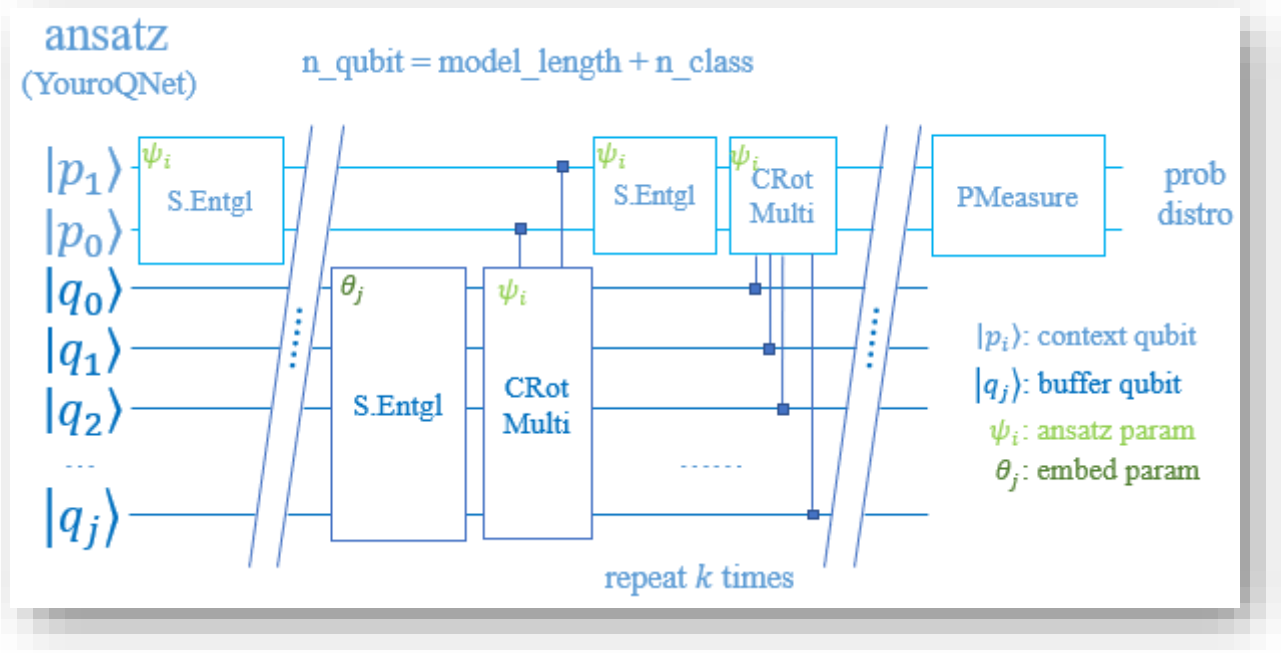


Content Table

YouroQNet for Quantum Sequence Classification

- YouroQNet :=
  - pure variational quantum circuit
  - representation learning
  - style-transfer scheme

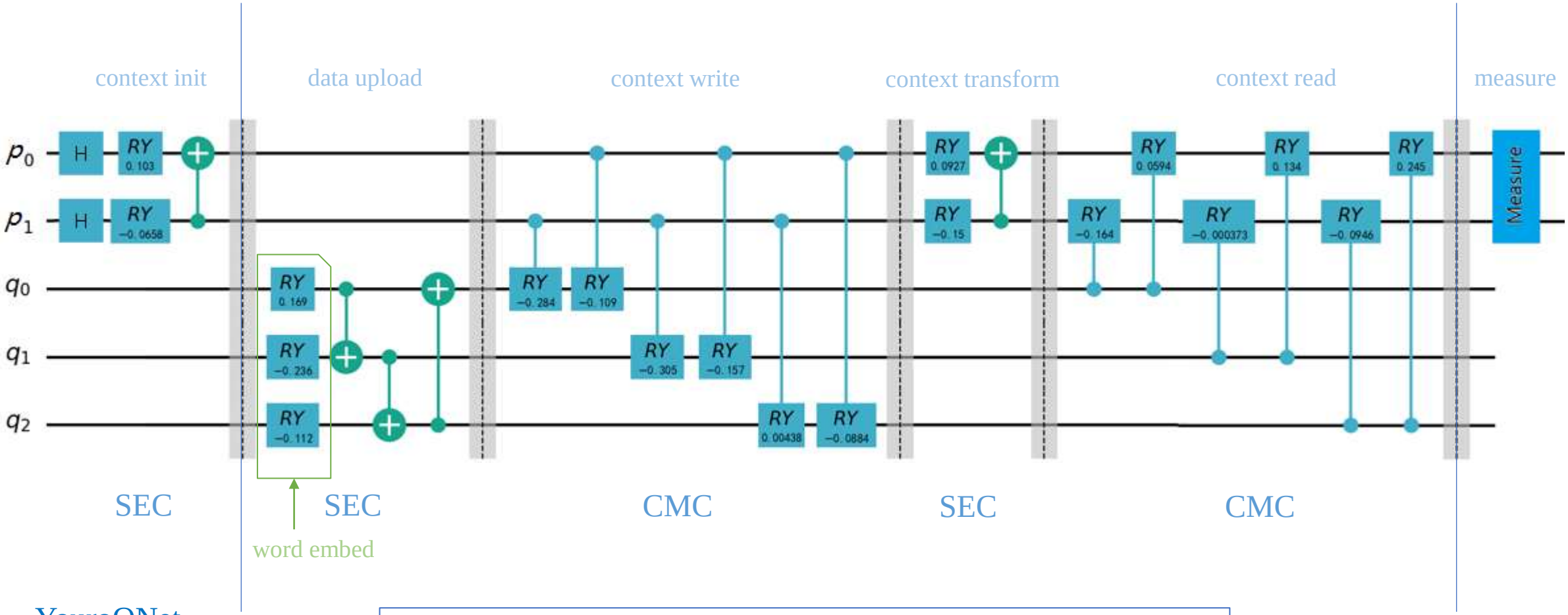
- what is for
  - sequence classification
  - feature abstraction



YouroQNet general architecture

YouroQNet: minimal concrete example

- pure variational quantum circuit



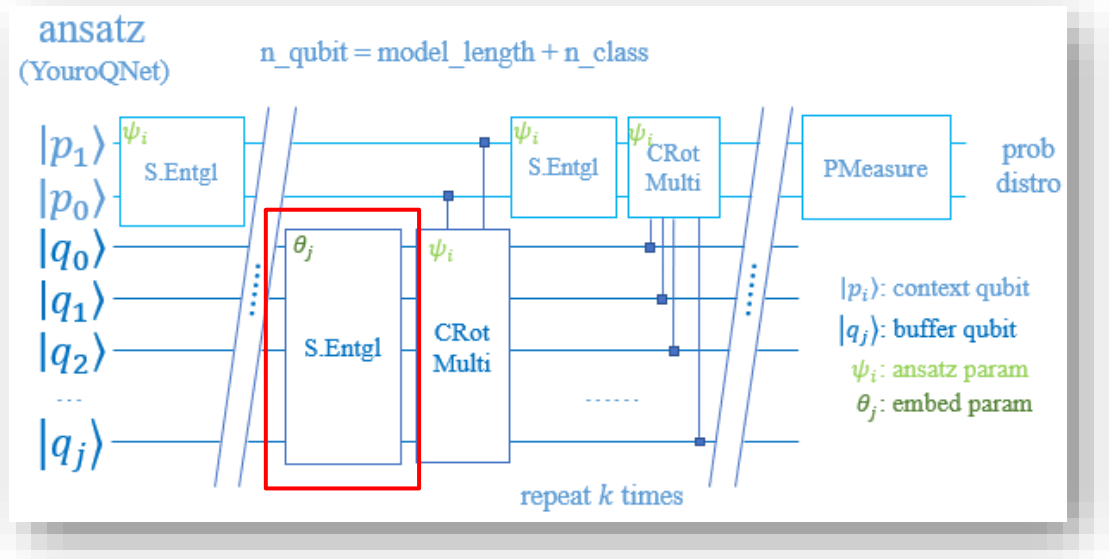
YouroQNet
(5qubit-1repeat)

SEC: Strong Entangle Circuit, here we set rots=[RY], entgl=CNOT
CMC: CRot Multi Circuit, here we set rots=[RY]

repeatable section

YouroQNet learns semantical word representation

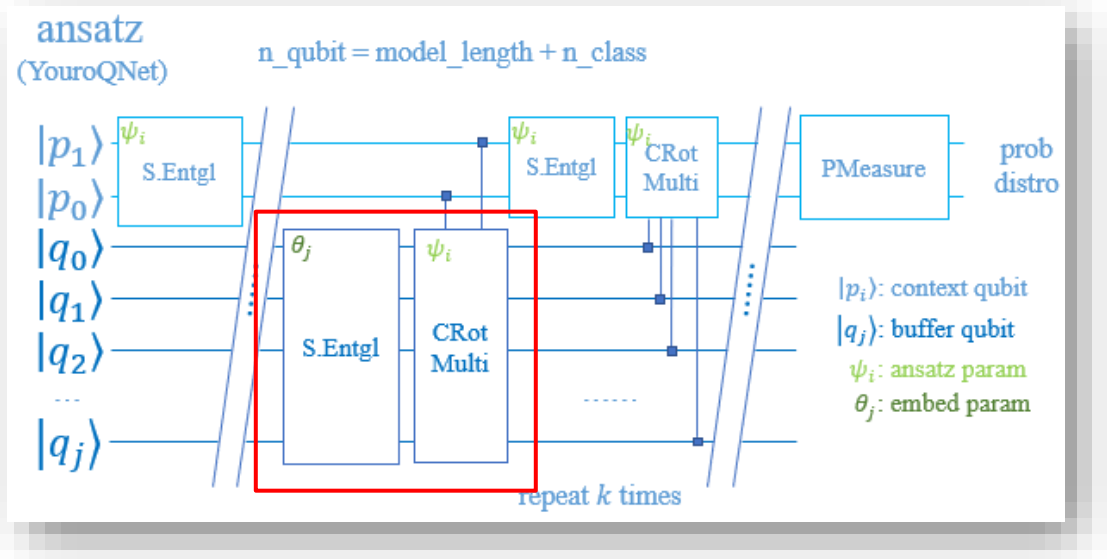
- YouroQNet :=
  - pure variational quantum circuit
  - representation learning
  - implicit convolutional
  - implicit recurrent
  - style-transfer scheme



Train the word embedding θ_j whose channels are partially temporally dependent.

YouroQNet differs contextual invariants from variants

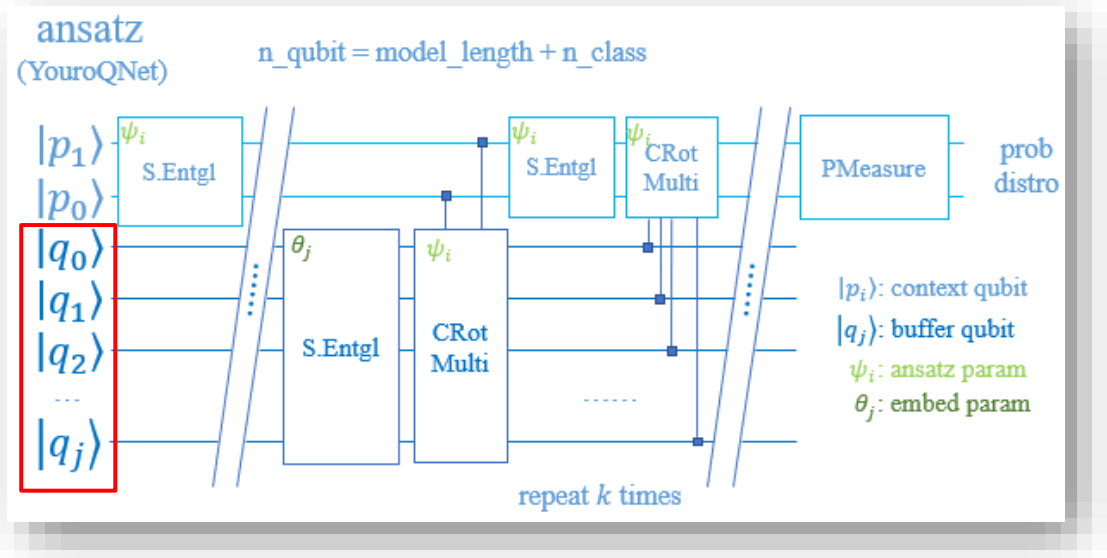
- YouroQNet :=
  - pure variational quantum circuit
  - representation learning
  - implicit convolutional
  - implicit recurrent
  - style-transfer scheme



Train ψ_i and θ_j alternatively alike the EM algorithm.
Note that ψ_i is constant to all inputs, corresponding to contextual invariant (syntactical) transforms, while θ_j builds up a (semantical) variant context.

YouroQNet is implicitly convolutional

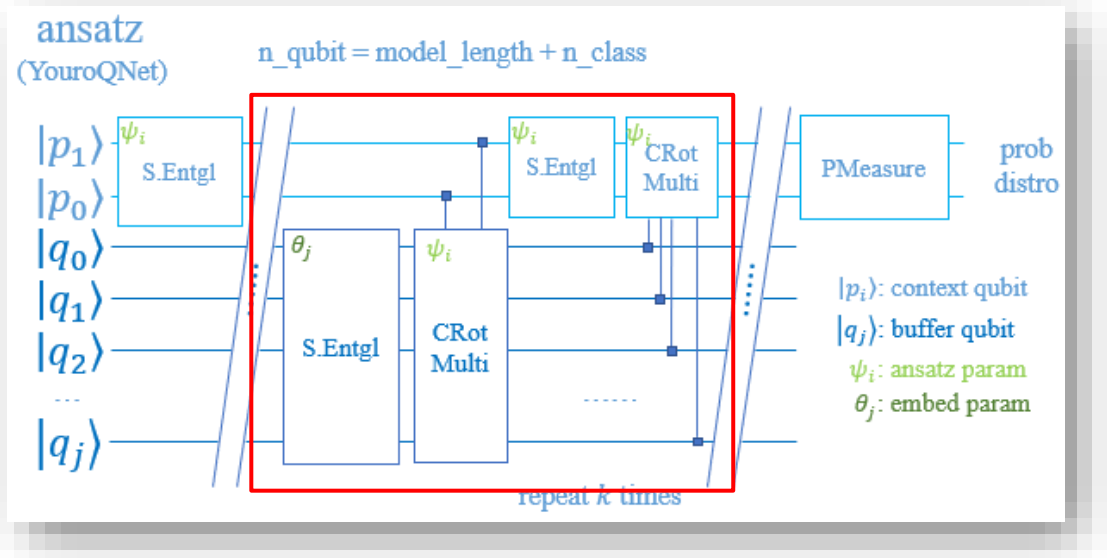
- YouroQNet :=
  - pure variational quantum circuit
  - representation learning
  - implicit convolutional
  - implicit recurrent
  - style-transfer scheme



Breaking down long sentences with the aligner, a YouroQNet is like a single Conv1d filter that slides along the sentence then applies an AvgPooling.

YouroQNet could be recurrent if needed

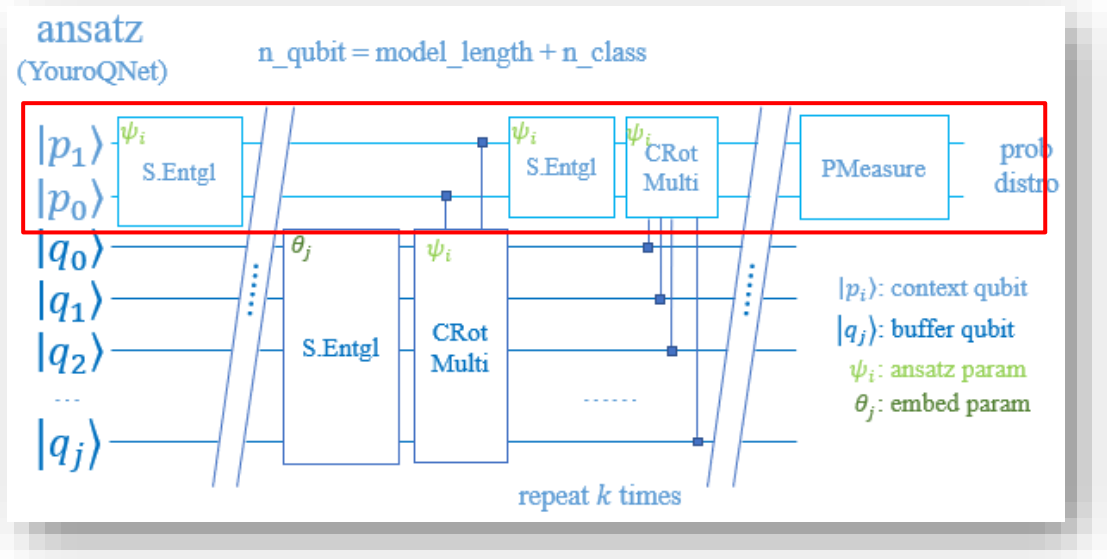
- YouroQNet :=
  - pure variational quantum circuit
  - representation learning
  - implicit convolutional
  - implicit recurrent
  - style-transfer scheme



The ancilla $|p_i\rangle$ is like a state memory, when ψ_i is shared across all repeatable parts of the circuit, it simulates an RNN.

YouroQNet follows a style-transfer scheme

- YouroQNet :=
  - pure variational quantum circuit
  - representation learning
  - implicit convolutional
  - implicit recurrent
  - style-transfer scheme



The ancilla $|p_i\rangle$ is the blank canvas, we gradually extract useful info from $|q_j\rangle$ and transfer onto it, through steps.

YouroQNet: metric evaluation & comparation

F1 score	Youro QNet	Text DNN	Text CNN	Text RNN	TF-IDF					FastText word2vec				
					kNN	GBDT	Bayes	SVM	MLP	kNN	GBDT	Bayes	SVM	MLP
Joy	0.018	0.349	0.321	0.287	0.338	0.214	0.357	0.374	0.41	0.361	0.413	0.430	0.458	0.455
Angry	0.342	0.294	0.175	0.32	0.327	0.356	0.300	0.336	0.291	0.347	0.333	0.375	0.393	0.370
Sad	0.22	0.306	0.321	0.283	0.175	0.169	0.274	0.349	0.306	0.310	0.24	0.171	0.328	0.225
Hate	0.273	0.187	0.275	0.318	0.271	0.259	0.308	0.24	0.313	0.179	0.289	0.306	0.314	0.290
Avg.	0.213	0.284	0.273	0.302	0.278	0.25	0.31	0.325	0.33	0.3	0.319	0.321	0.373	0.335

YouroQNet - mini: toy verification in details

- toy model config
  - 3+1 qubits, 1 repeat, [RY]-CNOT-[RY]
- embedding analysis
  - manually bias the dataset
  - embed highlights the minority
  - 👉 it learns TF-IDF successfully !!

```
words = {
  # positive (leading to class 0)
  '爱', '喜欢',
  # negative (leading to class 1)
  '恨', '讨厌',
  # neutral
  '啊', '我', '你', '苹果', '西瓜',
}
```

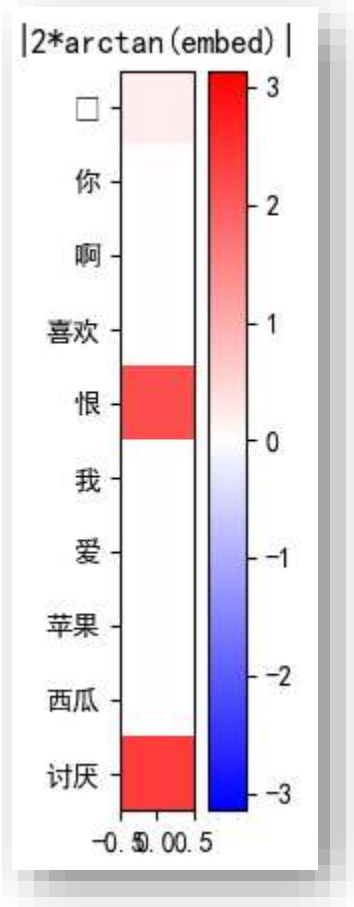
vocab

```
train_data = [
  (0, '我爱你'),
  (0, '我喜欢苹果'),
  (0, '苹果啊喜欢'),
  (0, '你爱西瓜'),
  (1, '你讨厌我'),
  (1, '讨厌西瓜苹果'),
  (1, '你讨厌苹果'),
  (1, '我恨啊恨'),
] + baised_data
```

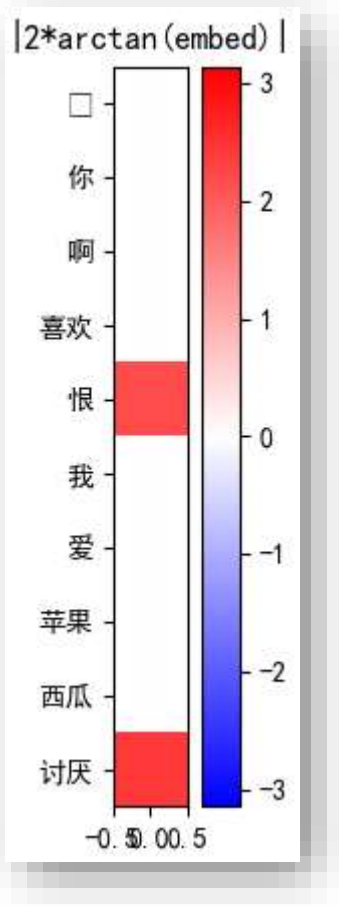
```
baised_data = [
  (0, '我你'),
  (0, '你啊'),
  (0, '啊我啊啊'),
  (0, '西瓜苹果'),
  (0, '你西瓜'),
  (0, '我苹果啊'),
]
```

pos = 0
neg = 1

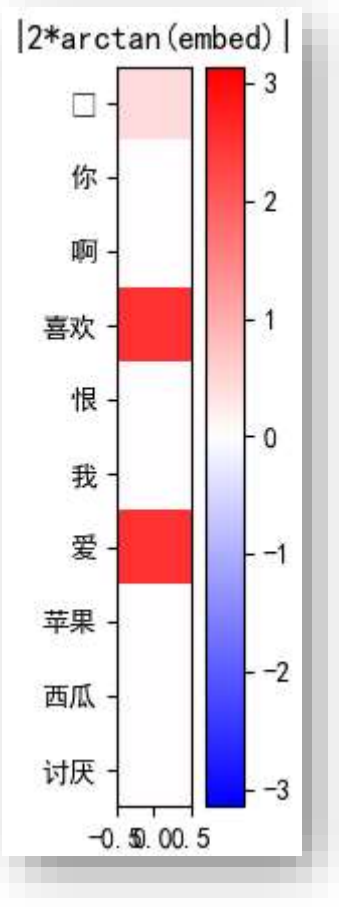
train data



neutral



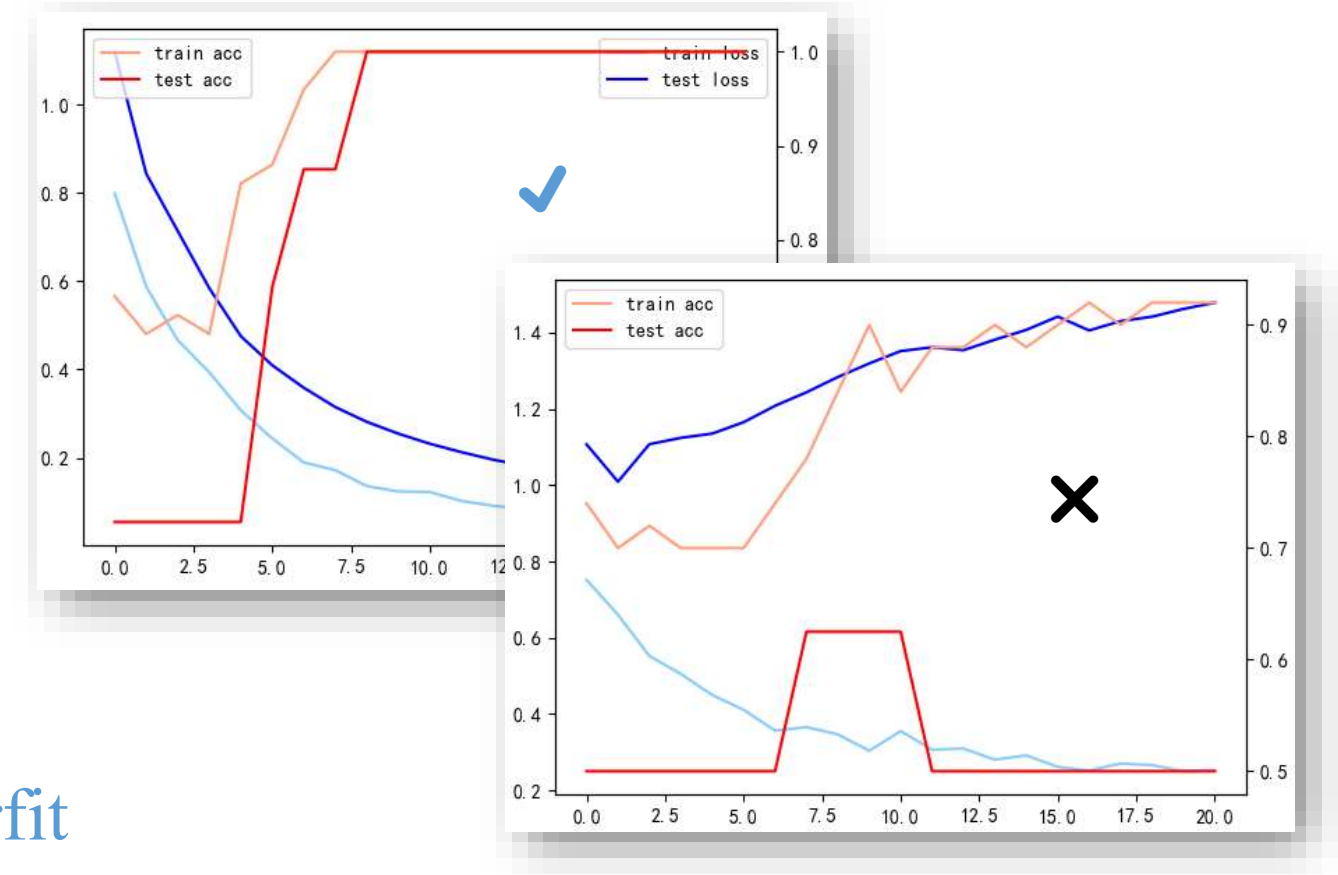
bias pos



bias neg

YouroQNet: QNN is fragile | symmetric, periodical, finite-valued

- parameter initialization
  - uniform ✗
  - normal ✓
- embedding normalization
  - ±pi/2 ✗
  - ±pi ✓
- gradient method
  - param_shift ✗
  - finite_diff ✓
- loss not decay or quickly overfit
  - tune rand seed
  - kill & retry 😅

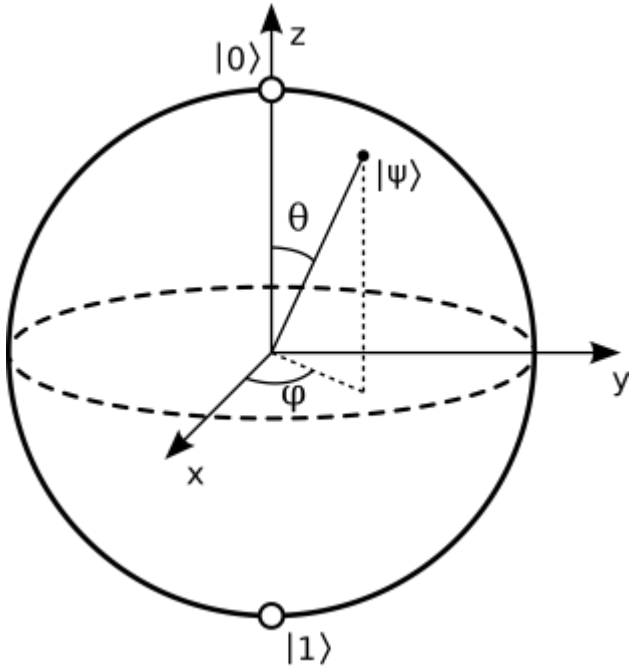


good & bad loss curves

Content Table

Rethink on QCricuit

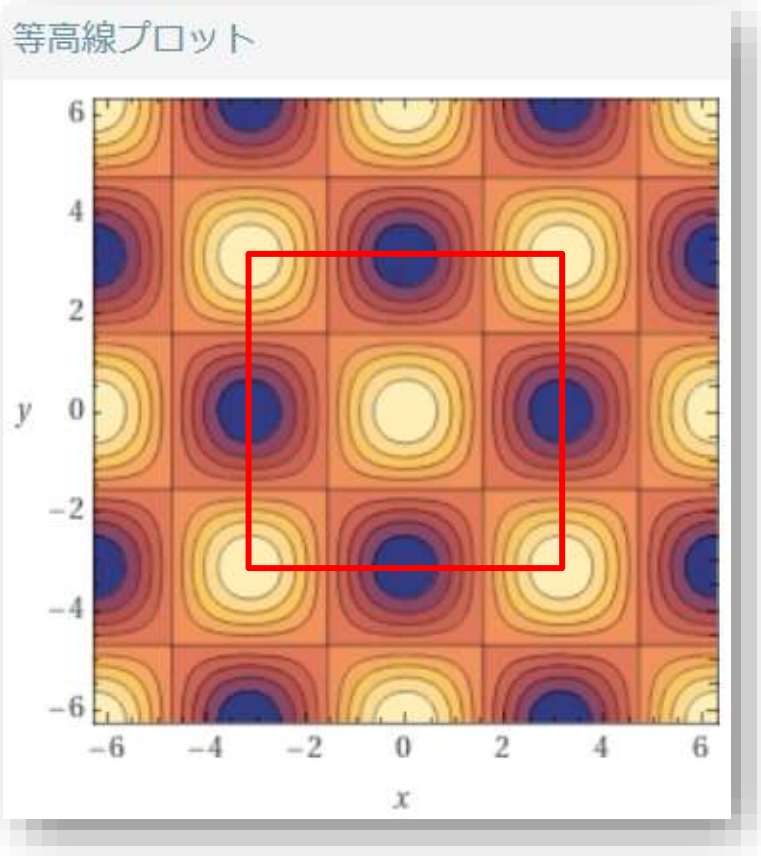
- Conceptual model in meme
  - Balancer / 3D Clock
  - Map-Reduce
- But what's the mathematical model?
  - Single qubit rotation (superposition)
  - Multi-qubit entanglement



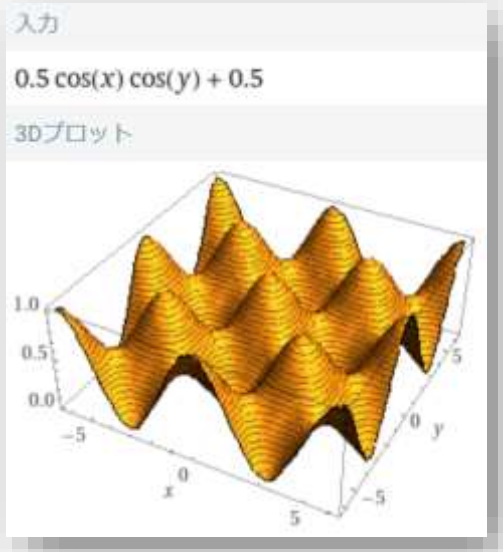
https://leimao.github.io/blog/Qubit-Bloch-Sphere/

XYZ - RotCiruit

- numerical XOR!!




```
# the single-qubit data encoder:
#  - 1 qubit
#  - 1~3 param
#  - rotate with RX/RY/RZ
#  - no entangle
# circuit:     c0          c1          c2
#
# q_0: |0>─┤RX(tht0)├─┤RY(tht1)├─┤RZ(tht2)├─
#
```

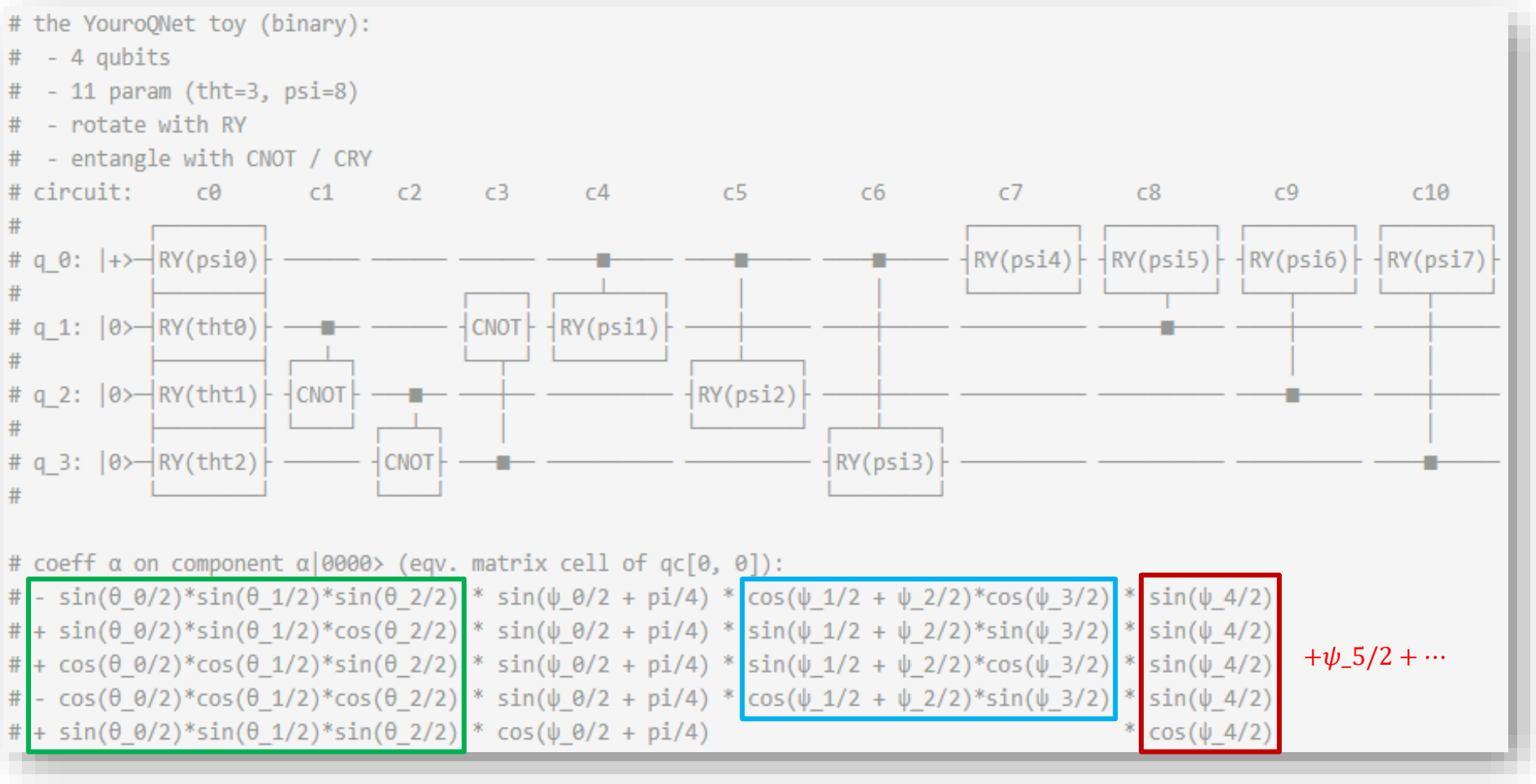



```
# coeffs on:
# |0>: ( I*sin(θ_0/2)*sin(θ_1/2) + cos(θ_0/2)*cos(θ_1/2))*exp(-I*θ_2/2)
# |1>: (-I*sin(θ_0/2)*cos(θ_1/2) + sin(θ_1/2)*cos(θ_0/2))*exp( I*θ_2/2)

# probs on:
# |0>:  0.5*cos(θ_0)*cos(θ_1) + 0.5
# |1>: -0.5*cos(θ_0)*cos(θ_1) + 0.5
```

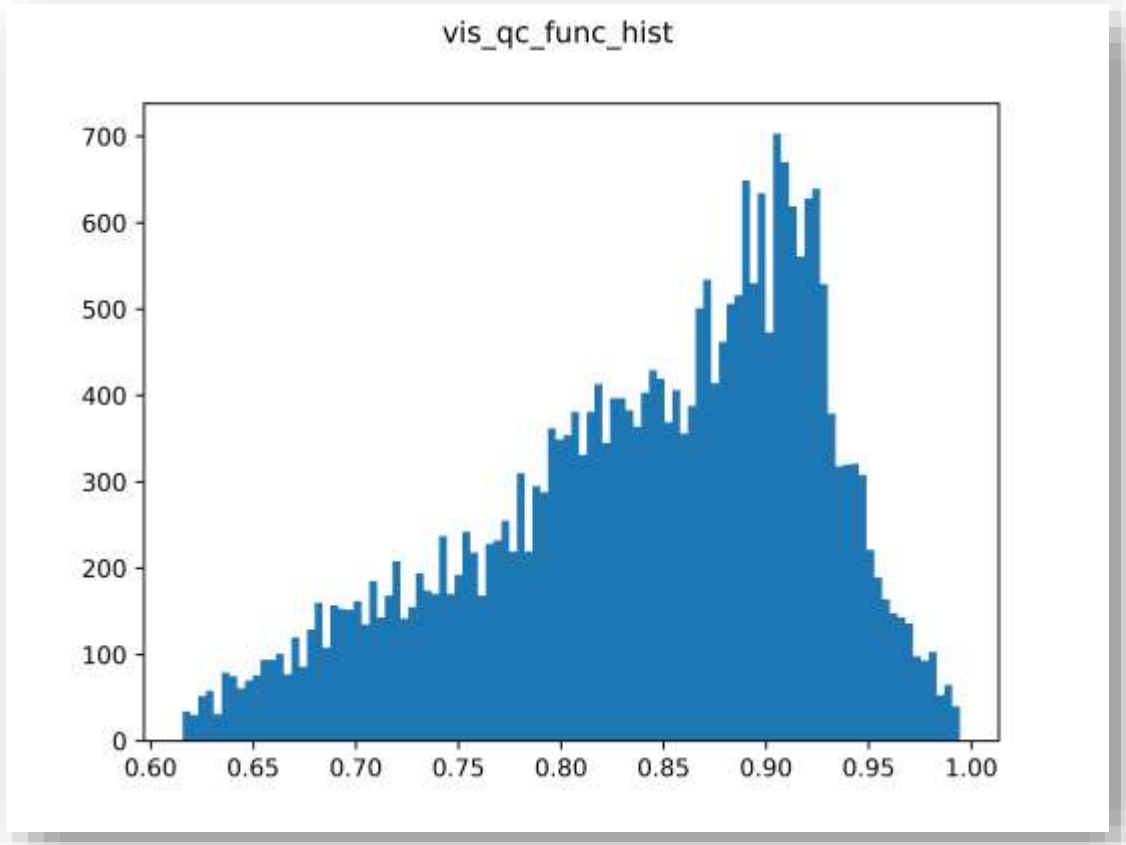
YouroQNet - mini

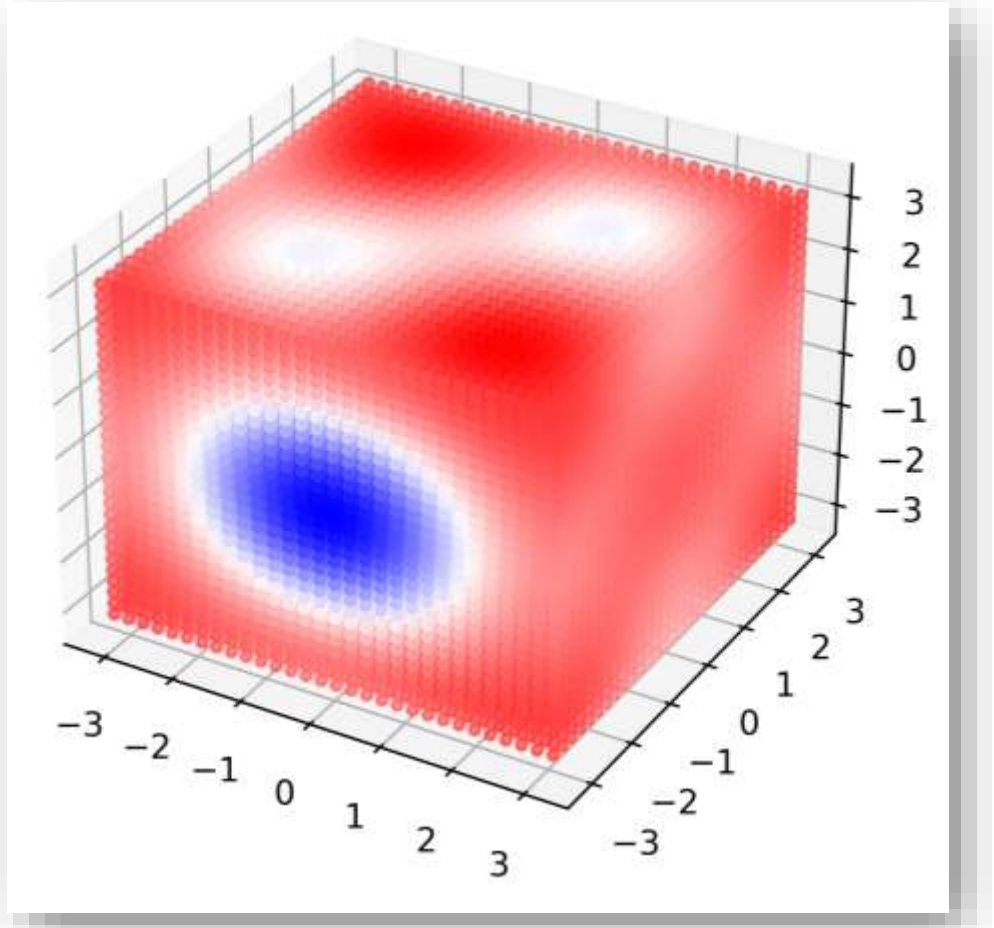
- combinational sum with trigonometric activation!!



YouroQNet - mini: learned function value space

- weird generalization…





Value space of a learned YouroQNet-mini
(input is word embedding θ)

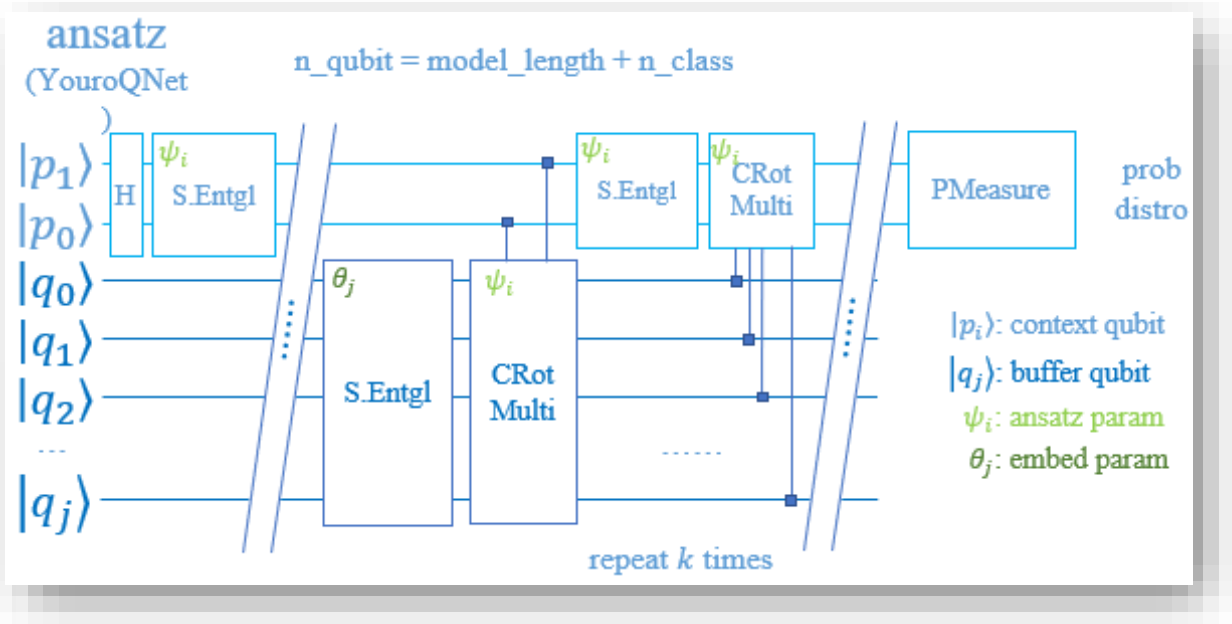
Content Table

- Entropy-based Heuristic k-gram Tokenizer
- YouroQNet for Quantum Sequence Classification
- Computational Analysis over Simple QCircuits

maximize $p(W) = \prod_i p(w_i)$

- p(这)p(是)p(一)p(个例)p(子)
- p(这是)p(一个)p(例)p(子)
- …

```
atrix cell of qc[0, 0]):
 sin(ψ_0/2 + pi/4) * cos(ψ_1/2 + ψ_2/2)*cos(ψ_3/2) * sin(ψ_4/2)
 sin(ψ_0/2 + pi/4) * sin(ψ_1/2 + ψ_2/2)*sin(ψ_3/2) * sin(ψ_4/2)
 sin(ψ_0/2 + pi/4) * sin(ψ_1/2 + ψ_2/2)*cos(ψ_3/2) * sin(ψ_4/2)
 sin(ψ_0/2 + pi/4) * cos(ψ_1/2 + ψ_2/2)*sin(ψ_3/2) * sin(ψ_4/2)
 cos(ψ_0/2 + pi/4)                                 * cos(ψ_4/2)
                                                                  +ψ_5/2 + ··
```



Thanks for your watching~

YouroQNet

熔炉ネットと言うのは、虚仮威し全て裏技を繋ぐもん

Team: QwQ

Reporter: Armit