**train\_final\_5**

text 100， 3 attention (10,20)

audio 25

fusion 100

text acc: 0.5376850605652759 audio acc: 0.38373655913978494 final acc: 0.5639300134589502

0.5338928343558634

0.5231513476775446

**1. attention**

1.1

**test\_1**

:text 100, 6 attention (8,64)

audio 0

fusion 0

text acc: 0.5282392026578073

0.5145255005536655

0.4914698162729659

1.2

**test\_1**

text 100, 3 attention (8,64)

audio 0

fusion 0

text acc 0.5171319796954315

text acc 0.5284900286598423

text acc 0.5386138613074526

1.3

**test\_1**

text 100, 3 attention (20,20)

audio 0

fusion 0

text acc 0.5206942593303319

text acc 0.5202702702702703

**2.audio**

**Test\_1**

audio 13, 3 attention (4,16),no dropout, globalmaxpooling, adam

text 0

fusion 0

model

acc\_a 0.3554987212276215(模型不在了)

0.37890625

audio 13, 3 attention (4,16),no dropout, globalmaxpooling, sgd

text 0

fusion 0

model\_ testaudo

0.45866141732283466

**3. adma + sgd**

**3.1**

**Test\_2**

text 100, 3 attention (10,20),no dropout, globalmaxpooling, adam

audio 0

fusion 0

text acc 0.5408031088082902

text acc 0.5202429153015096

model

text 100, 3 attention (10,20),no dropout, globalmaxpooling, sgd

audio 0

fusion 0

text acc 0.7973258275536061

text acc 0.8057455540355677

text acc 0.8021319126939869

model1

3.2

Test\_2

text 50, 3 attention (10,20),no dropout, globalmaxpooling, adam

audio 0

fusion 0

text acc 0.47956403302236866

text 50, 3 attention (10,20),no dropout, globalmaxpooling, sgd

text acc 0.7904761908816642

model2

3.3

Test\_2

text 100, 3 attention (10,20),no dropout, globalmaxpooling, adam

audio 0

fusion 0

text acc 0.5269102990033223

model3

text 100, 3 attention (10,20),no dropout, globalmaxpooling, sgd, model3

audio 1, ->audio\_model2

fusion 100， ->final\_model2,

text acc 0.7456197274690476

text acc: 0.768135593220339 audio acc: 0.3353658536585366 final acc: 0.783728813559322

0 {'0': 147, '1': 7, '2': 2, '3': 56, '4': 11}

1 {'0': 3, '1': 283, '2': 2, '3': 6, '4': 28}

2 {'0': 0, '1': 5, '2': 178, '3': 15, '4': 16}

3 {'0': 19, '1': 17, '2': 9, '3': 295, '4': 45}

4 {'0': 4, '1': 26, '2': 18, '3': 30, '4': 253}

text acc: 0.7639257294429708 audio acc: 0.34748010610079577 final acc: 0.7844827586206896

0 {'0': 185, '1': 14, '2': 3, '3': 36, '4': 16}

1 {'0': 1, '1': 274, '2': 4, '3': 6, '4': 30}

2 {'0': 1, '1': 5, '2': 177, '3': 14, '4': 15}

3 {'0': 35, '1': 12, '2': 14, '3': 272, '4': 38}

4 {'0': 6, '1': 26, '2': 20, '3': 29, '4': 275}

**4.fusion without cnn**

**File:Test\_2**

T1:

text (100,100) 3 attention (10,20),no dropout, globalmaxpooling, adam,sgd, text\_model3

audio 1 , audio\_model2,

fusion 100, no cnn ->final\_model2

text acc: 0.7768651608691324 audio acc: 0.34084699453551914 final acc: 0.7967145790758401

0 {'0': 161, '1': 3, '2': 2, '3': 29, '4': 20}

1 {'0': 2, '1': 298, '2': 2, '3': 11, '4': 23}

2 {'0': 0, '1': 9, '2': 179, '3': 5, '4': 15}

3 {'0': 31, '1': 8, '2': 18, '3': 260, '4': 37}

4 {'0': 3, '1': 26, '2': 20, '3': 33, '4': 266}

T2:

text (100,100) 3 attention (10,20),no dropout, globalmaxpooling, adam,sgd, text\_model3

audio 1 , audio\_model2,

fusion 100, no cnn ->final\_model2

text acc: 0.7711864411442345 audio acc: 0.3365885416666667 final acc: 0.796610169802371

0 {'0': 164, '1': 7, '2': 1, '3': 39, '4': 12}

1 {'0': 2, '1': 311, '2': 5, '3': 5, '4': 25}

2 {'0': 1, '1': 7, '2': 185, '3': 22, '4': 12}

3 {'0': 18, '1': 7, '2': 13, '3': 303, '4': 28}

4 {'0': 7, '1': 23, '2': 25, '3': 53, '4': 259}

5.position\_embedding

Only text branch 50 with adam without pe with Maxpooling

0.5401907358572333

0.5444444444444444

0.5374746794317479

6.max or average

Only text branch 50 with adam with pe

Max: 1. 0.5380039656508938

2. 0.5371269951422624

3. 0.5239030780086504

Average: 1. 0.520958084070278

2. 0.5333773959218786

3. 0.523346303656562

6. 4-class

File: Test\_4\_class, test\_load\_4\_class

Max adam 0.6409749675331066

Text ,100(adam), 100(sgd), max-> text\_model\_4\_class.h5

1. Adma 0.6398362895476379
2. Sgd

text acc: 0.80234163918374621 audio acc: 0.4728835978835979 final acc: 0.8558201058201058

0 {'0': 192, '1': 2, '2': 2, '3': 37}

1 {'0': 2, '1': 281, '2': 12, '3': 33}

2 {'0': 2, '1': 3, '2': 185, '3': 26}

3 {'0': 31, '1': 32, '2': 36, '3': 636}

Audio ,1,

Fusion ,100, no cnn

**7.fusion without cnn, with sgd**

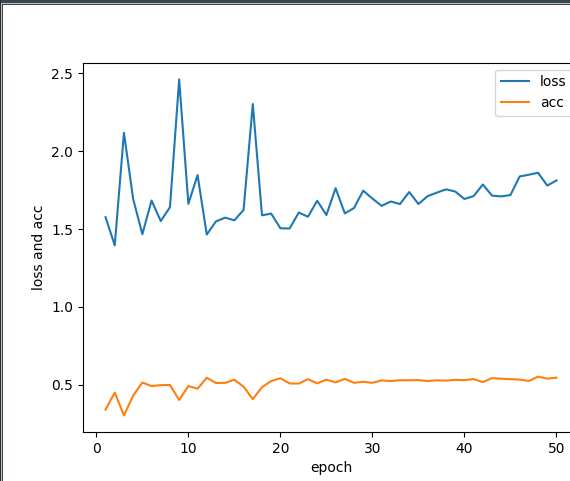
8. without cnn we can use different att as the output

9.dense self -attention

1/27

1. Plot audio loss curve

Text 50 epoch(adm)



loss: [1.5763737176661536, 1.3955214256232713, 2.118301702540463, 1.693521633443409, 1.4665021963748444, 1.6833600924120937, 1.5511946042806593, 1.640583092158768, 2.461524157312324, 1.6619793155793385, 1.846784676709695, 1.4644518298865328, 1.549314581850496, 1.573477045844734, 1.555990247193769, 1.6228457307430488, 2.303072814825887, 1.588209819986233, 1.5994921576126426, 1.5051681146647375, 1.503670638534936, 1.6059443308078103, 1.5792582485910058, 1.6820010535649625, 1.5901791327413608, 1.762555205164174, 1.6006172568127413, 1.6349813925305414, 1.7468356583994458, 1.6958976627840168, 1.6490668940832927, 1.6771894662890428, 1.6605865318393322, 1.7382027005443341, 1.660755176884329, 1.7114445928930435, 1.7340766669603247, 1.7550948672147009, 1.7412151661406936, 1.693297959112062, 1.711881248979838, 1.786170222075759, 1.7148854506288531, 1.70990935838399, 1.7186814986199421, 1.8379550151837787, 1.8489847639688413, 1.8618124051177485, 1.7794996113347172, 1.81229015891998]

acc [0.340511440127727, 0.44952893682315564, 0.30282637962261727, 0.42934051148021846, 0.5141318979526432, 0.49125168240888606, 0.49730820995962316, 0.4986541052204764, 0.4017496635463004, 0.49125168239885836, 0.47442799461641993, 0.5450874830960907, 0.510767160241729, 0.5121130552619174, 0.532301480725187, 0.48721399754887473, 0.40713324364710946, 0.4838492597978497, 0.5235531628532974, 0.5417227457060628, 0.5080753703617953, 0.5074024224505932, 0.5356662182756581, 0.5087483175510031, 0.5323014804845222, 0.5148048452621835, 0.5376850607257191, 0.5121130552218066, 0.5188425303227487, 0.5121130551014742, 0.5275908481545275, 0.5235531630939623, 0.528936743174716, 0.5282637953437354, 0.5296096903639238, 0.5235531628532974, 0.528263795423957, 0.5262449530541174, 0.5316285330145388, 0.528936743174716, 0.5363391655851982, 0.5168236879930201, 0.5423956932963786, 0.5383580079951485, 0.5356662181152149, 0.5329744282352812, 0.5242261105238348, 0.5518169585179191, 0.5390309558261291, 0.5450874830960907]

predict : (evaluate bias)

loss [0.6155164779729585]

acc [0.8215465962059515]

Text 50 epoch(sgd):

Load\_weights evaluate bias:

Text 50 (adam), audio 1 (adam), fusion 10 (adam):

final result:

text acc: 0.5434491978609626 audio acc: 0.32085561497326204 final acc: 0.3409090909090909

0 {'0': 18, '1': 106, '2': 9, '3': 31, '4': 42}

1 {'0': 22, '1': 147, '2': 14, '3': 63, '4': 83}

2 {'0': 0, '1': 6, '2': 106, '3': 16, '4': 88}

3 {'0': 5, '1': 152, '2': 42, '3': 89, '4': 136}

4 {'0': 7, '1': 62, '2': 45, '3': 57, '4': 150}

final result:

text acc: 0.5463087248322148

0 {'0': 34, '1': 55, '2': 26, '3': 38, '4': 59}

1 {'0': 50, '1': 81, '2': 43, '3': 64, '4': 108}

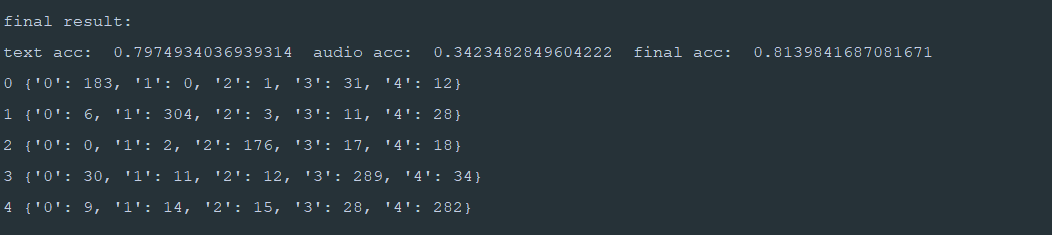
2 {'0': 32, '1': 62, '2': 30, '3': 42, '4': 54}

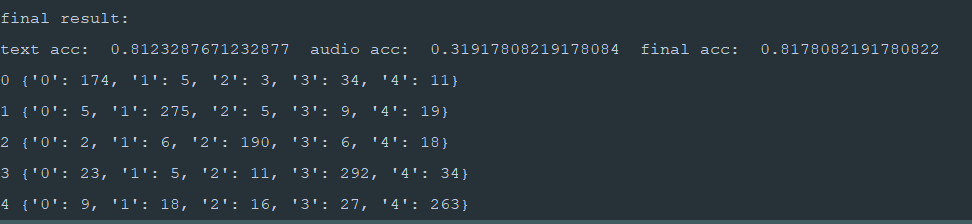
3 {'0': 65, '1': 101, '2': 58, '3': 59, '4': 89}

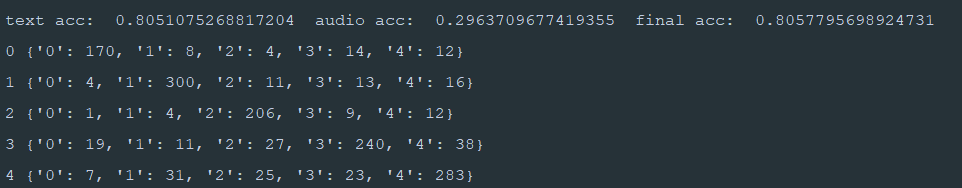
4 {'0': 47, '1': 81, '2': 58, '3': 65, '4': 89}

Load Text 50 (adam), audio 1 (adam), fusion 10 (adam):

Test\_6:







Time: audio train 1 epoch 6min10s