Research in Lin-Shan's Lab 解正平、吳倉永

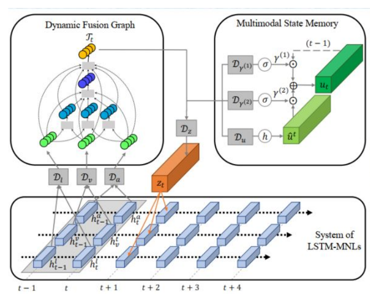
❑ Research Motivation

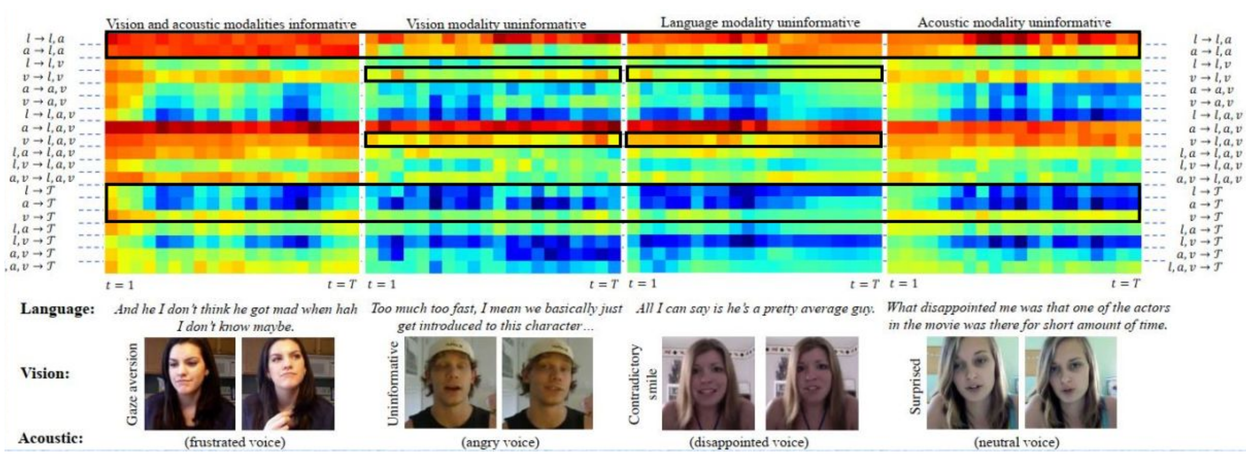
There are lots of style transfer research in text. We look forward to building an emotional chatbot which can have different emotion in conversation. Not only we can just change the response in emotional text style but also can generate a style text by setting emotion in advance. However, to achieve our goal, we have to learn a machine to classify emotion which is a difficult problem nowadays. Therefore, we devoted ourselves to these problems and will go on until receiving good result.

❑ What We have Learned

- **10/1-Multimodal Emotion and Sentiment Classification**

  Recently, the emotion and sentiment classification problem is regarded as a multimodal problem. There is more and more research devoting on the relationship between different modalities, especially text, video and audio. We can use early fusion, late fusion, hierarchical fusion or tensor fusion network[1] to implement different architectures for learning modality relationship. The most impressive work is that Zadeh et al[2] use memory fusion network to get the dynamic fusion feature which can let us know how the deep neural network had learned modality information. The importance between text, video and audio can be visualize on the weighted parameters as followed.





- **10/1-Emotional Chatting Machine**

  Studied on Chinese Emotional Chatting Machine capable of generating emotional sentences. It utilized three methods to control the emotion state in RNN-seq2seq model.

  1. Emotion category embedding: use vector to represent an emotion
  2. Internal emotion memory: balance between grammar state and emotion state
  3. External emotion memory: choose emotion or generic vocabulary

Result of emotional accuracy are presented below. ECM outperformed baseline seq2seq model and if all of the techniques are utilized, the performance would be much greater.

Method	Perplexity	Accuracy
Seq2Seq	68.0	0.179
Emb	62.5	0.724
ECM	65.9	**0.773**
w/o Emb	66.1	0.753
w/o IMem	66.7	0.749
w/o EMem	**61.8**	0.731

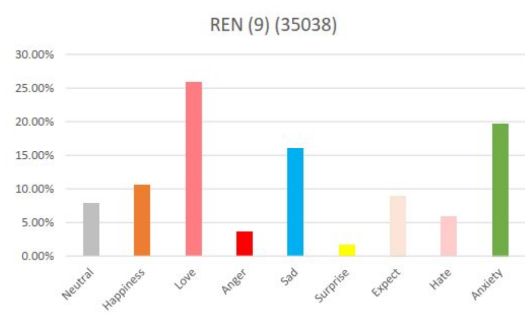
- **10/8-Dataset**

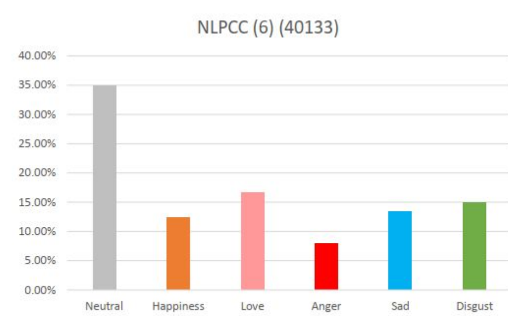
  We collect some different dataset[3] for our emotion dialogues task. There are prepared datasets including dialogues corpus with emotion label and the sentence with emotional label which can build an emotion classifier to label a large scale corpus dataset. We have datasets in Chinese or English and we choose Chinese for our research topic.
  - Emotional Dialogues Corpus Dataset (Chinese and English)
  - Emotional Classifier Dataset (Chinese and English)
  - Dialogues Corpus Dataset (Chinese and English)

- **10/22-Data Analysis**

  We use REN(REN CECps 1.0)[4] for our major dataset and NLPCC(2014 NLPCC)[5] for our auxiliary dataset. REN contains 35038 sentences with 8 emotion labels and NLPCC contains 40133 sentences with 6 labels. REN have more balanced label distribution while NLPCC have a dominated Neutral label in 35%. REN have label intensity which indicates that we can have multi-label and distribution in one sentence while NLPCC only have one emotion label in one sentence.

  We use the pre-trained dictionary[6] from Beijing Normal University as our word embedding





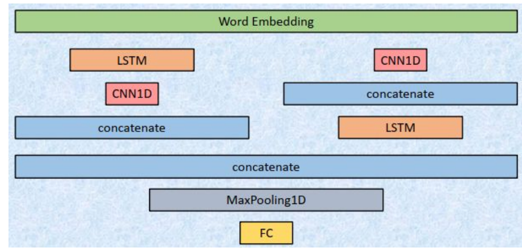
which is the state of the art in chinese word representation. Different from Facebook Fast Text training on Wikipedia, the word embedding we use had trained on various daily news, books, social media, and encyclopedia.

- **11/19-Fundamental Training Strategy**

  We implement different model architecture to improve our performance and also carry out some tips to solve the classification problem. There are some conclusions in our experiments.

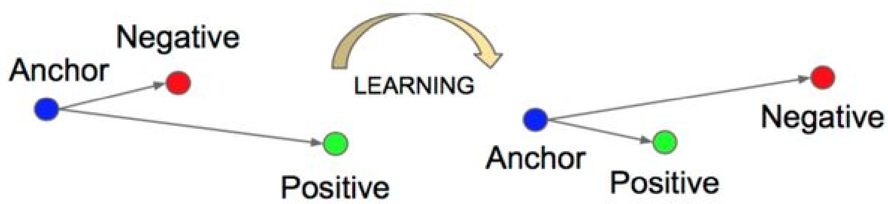
  - CNN and RNN model

    First, we have bilateral RNN model as our baseline with 46% accuracy on REN. Also, We try to use CNN 1D model know whether it can improve the performance. However, it doesn't affect the result. Furthermore, we fix or train the embedding layer which still have little influence. As a result, we have a conclusion that more parameters or deeper architecture enhance little.



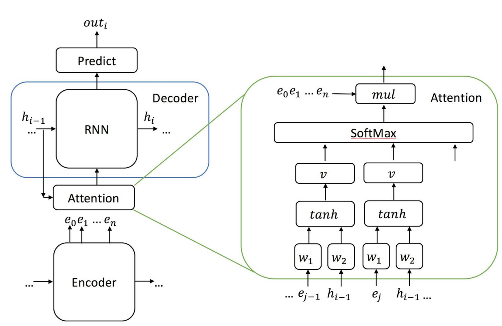
  - Metric Learning model

    To improve performance on classification, metric learning can really help to solve problems. Siamese and Triplet Network are the common methods to split different class data. For example, the same class data need as closer as much while different class really need far and far. Therefore, we control the feature distance on the space and try different distance functions. However, it doesn't have effective help on emotion task due to the class relationship which means not all class is independent and need a distance to split. As a result, we think we have to figure out other methods to fix this problems.



  - Attention based model

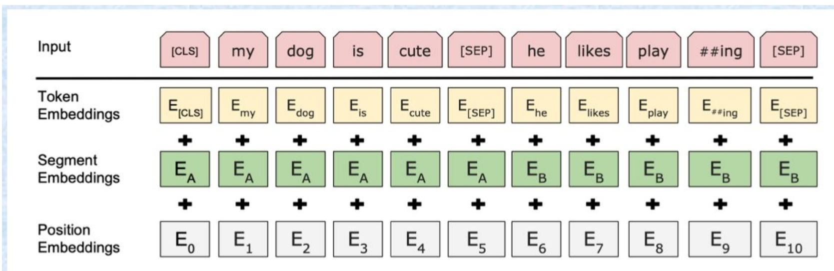
    In the past research, attention-based model can really work in the word representation task. We implement two attention model, one is attention with the last layer to pay more emphasis on the important feature; another is self attention with the feature itself not the last timestamp feature of RNN model. The results for attention based model are 48% and 50% which is really helpful for our task slightly.



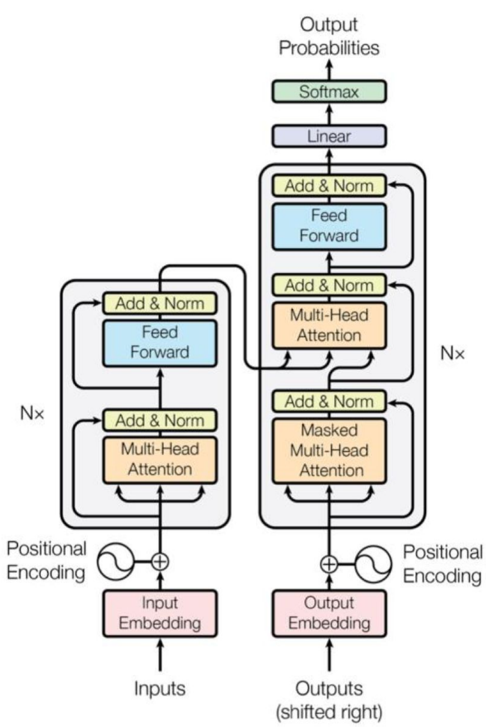
○ BERT and Transformer

BERT, Bidirectional Encoder Representations from Transformers, is proposed by Google. And it is the state-of-the-art model that could learn position-related information that outperformed RNN on several text-based tasks and all we have to do is to fine-tune BERT on the tasks you would like to solve. And Transformers is what makes bert special.

1. **Transformer**
    a. Based solely on attention Mechanism
    b. Position embedding
    c. Learn position-related information globally

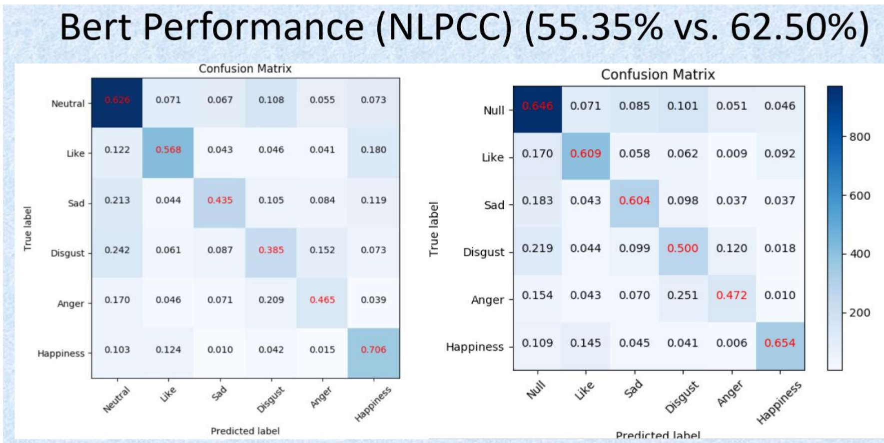


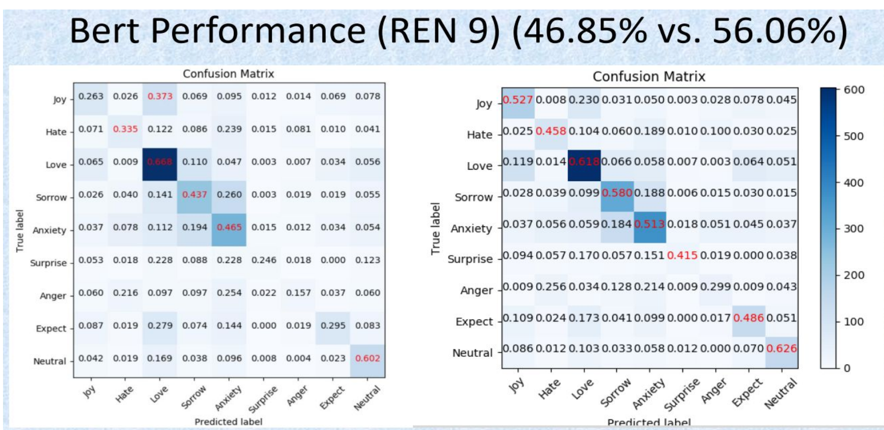
2. **BERT Hyper-Parameters Tuning**
    a. Batch size: 16, 32
    b. Learning rate: 5e-5, 3e-5, 2e-5
    c. Number of epochs: 3, 4



- **12/10-Performance Analysis**
  - Confusion matrix

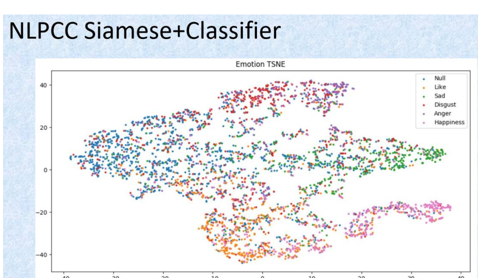
    We can find that BERT performance really defeat the RNN performance which indicates that attention is what we need. Most importantly, BERT has little influence by the imbalanced dataset. However, we still need to carry out a method to deal with the class relationship problem, such as ‘happy, expect, like’ and ‘sorrow, anxiety’ and ‘anger, disgust’. The inter-class problem leads to hard classification in emotion.

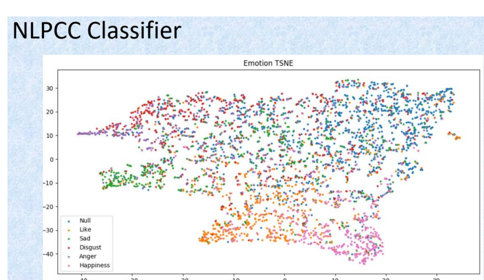




  - TSNE

    In the TSNE image, we can find the classifier can really split the class distribution and siamese can collect the same class data closely without higher performance. The anger(purple) and disgust(red) or the neutral all mixed together in multiple classes which really the problem is.





- Sentence Testing

  To compared our emotion classifiers performance between BERT and RNN modeling, we tried some confused input sentence that have two opposite emotion words in a sentence. Like the one below

	BiRNN			BERT		
今天天氣不錯，但是心情很糟	Happiness	Joy	Like	Anger	Anxiety	Sad

  The above sentence has positive emotion “不錯" and negative "很糟". However, the whole sentence would be negative emotion. As you can see,m BiRNN modeling could not correctly predict its emotion, and turned out to be "Happiness" or "Joy". But BERT modeling is capable of distinguish emotions from them, and find the most critical and influential parts in the sentence, which is "但是心情很糟".

  It shows how good are the BERT representations to separate different emotional words and how great is BERT modeling to learn position-related information from a sentence to eliminate positive emotion.

- **12/31-Enhanced Training Strategy**

  There are lots of methods to deal with the emotion class relationship problem. And We carry out the following idea including single or multiple labels and hope to solve the problem.

  - Single Label

    We can see our problem as single label. In one sentence, we get the highest probability or the highest intensity as our label. The training method include metric learning, class distance learning, kl distance learning. The all methods implement the concept of relationship between classes and help learn cross entropy loss effectively.

    Label：Get the highest probability or get the highest intensity label

    Method：metric learning, class distance learning, kl distance learning

    Evaluation：Accuracy

  - Multi Label

    There may be multiple emotions In one sentence; therefore, we get multi label with intensity or ones and zeros in our task. In ones and zeros, we have to give a threshold to label the sentence and train the model with multi-classifier which is multiple binary cross entropy task. In intensity, we can train the model as a multiple regression problem or still use multiple binary cross entropy to learn distribution for compare. Moreover, in multi-label problem, we can implement GAN to let generator learn as a classifier and discriminator to determine whether its a good label. We think it is a try to give some experiments on this problem.

    Ones and zeros Label：use threshold(0.5) to get ones and zeros label

    Ones and zeros Method：multiple binary cross entropy loss or GAN

    Ones and zeros Evaluation：F1 / Accuracy with the highest probability label

    Intensity Label：use intensity to all label class

    Intensity Method：multiple binary cross entropy loss or multiple MSE loss or GAN

    Intensity Evaluation：MSE / F1 threshold label / Accuracy with the highest probability label

❑ Future Work

We have two choices to do in the future

1. Emotional Classifier Modeling
2. Style Transfer Modeling

First, Emotional Classifier Modeling. In this task, we can use multi-labels and revised loss function to train a more robust classifier that could outperform current single-labeled modeling. Use F1 score or other performance metrics to compare the performance between. Finally, we may be able to proposed a BERT-revised modeling that dominate the emotional classifier performance in current framework.

Second, we can use current BERT-based classifier to later combined with GAN to propose a Style Transfer Modeling with good transfering results. And this is more like a application-based topic, so it the results would be much more stunning and fun.

In the end, to choose our topic, we may first try to reproduce the paper proposed by Professor's lab in April 2018, which was a persona style transfering, but it was only 2 emotions, on ours work , which would be 6 ~ 9 labels to see if we could do some analyze on the modeling. On the other hand, we would work on BERT and train a multi-labeled modeling with the methods described in previous section to make some breakthrough. **To sum up, we would try both of them on accessible materials, wishing to have inspiration on us.**

❑ Reference

[1] Tian, Leimin, Catherine Lai, and Johanna D. Moore. "Polarity and Intensity: the Two Aspects of Sentiment Analysis." *arXiv preprint arXiv:1807.01466* (2018).

[2] Zadeh, AmirAli Bagher, et al. "Multimodal language analysis in the wild: Cmu-mosei dataset and interpretable dynamic fusion graph." *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*. Vol. 1. 2018.

[3] Dataset Collect on our own：https://goo.gl/vRm9zF

[4] Quan, Changqin, and Fuji Ren. "Construction of a blog emotion corpus for Chinese emotional expression analysis." *Proceedings of the 2009 Conference on Empirical Methods in Natural Language Processing: Volume 3-Volume 3*. Association for Computational Linguistics, 2009.

[5] NLPCC Emotion Analysis in Weibo Texts http://tcci.ccf.org.cn/conference/2014/dldoc/evatask1.pdf

[6] Chinese Word Vectors https://github.com/Embedding/Chinese-Word-Vectors

Research in Lin-Shan's Lab

解正平

❑ Research Motivation

There are lots of style transfer research in text. I look forward to building an emotional chatbot which can have different emotions in conversation. Not only we can just change the response in emotional text style but also can generate a style text by setting emotion in advance. In this semester, I research on some related works and have a few results of the emotional chatbot.

❑ What We have Learned

- **Text Style Transfer Overview**

Nowadays, there are lots of methods for the text style transfer, especially on the emotional positive and negative. Someone had used multi-decoder for each style and design autoencoded and an encode classifier to make the content representation rid of style. (Fig.1) Another work is that they use cross-aligned autoencoder with GAN based adversarial loss to project generated and true samples have similar distribution. (Fig.2)

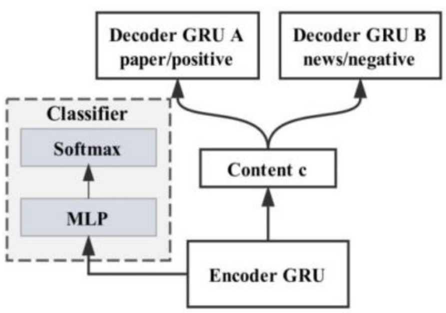


Fig.1

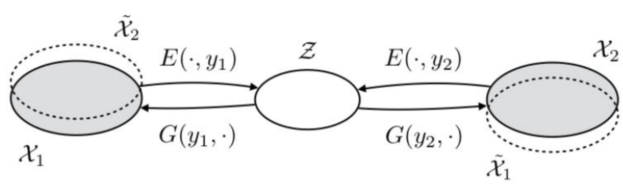


Fig.2

More works had considered GAN based methods for style transfer. For example, we can use an architecture like infoGAN. (Fig.3) First train a VAE model for unlabeled samples and train a generator and discriminator with labeled samples. The latent code z can have a constraint to reconstruct with Encoder while c can have a constraint to reconstruct with Discriminators. Besides the GAN based concepts on style and content, we can also use language model as a discriminator to force the generator give sentence like true sentence. (Fig.4)

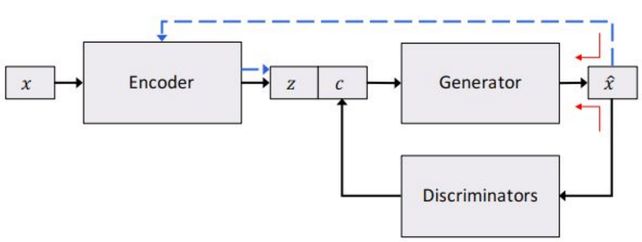


Fig.3

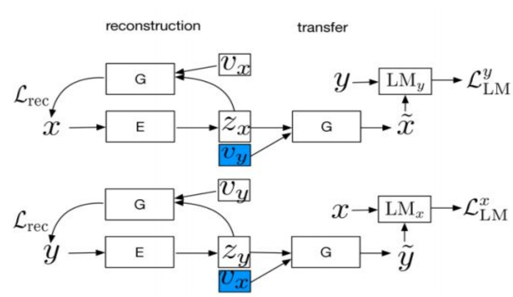
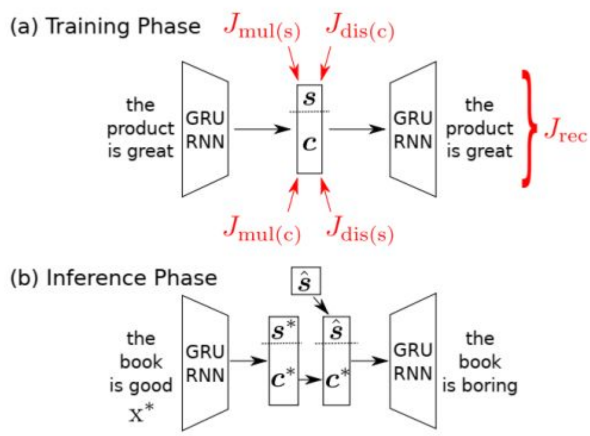


Fig.4

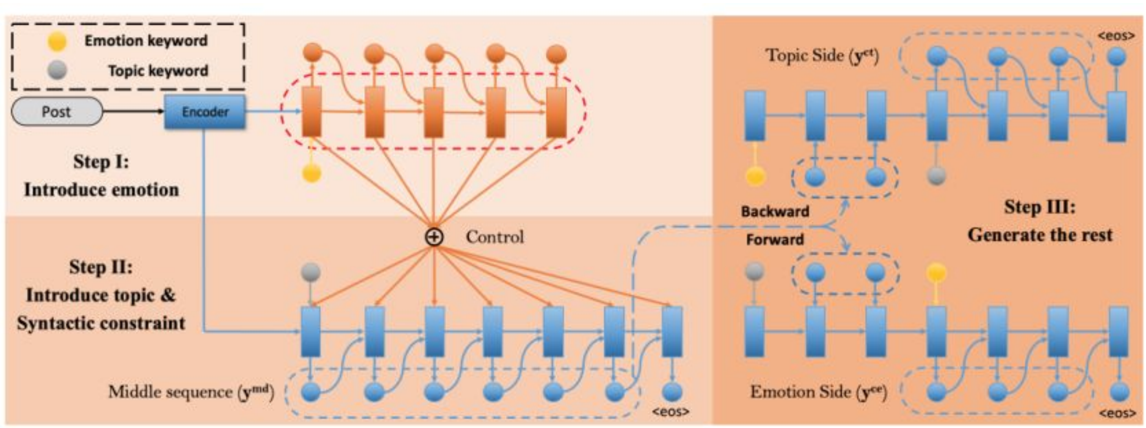
With disentangled methods, there is a research working on the disentangle of content and style by building a multi-task loss classifier and an adversarial loss classifier. The style vector is represented by labels and content vector is represented by BoW features, just like Fig.5.



There are some difficulties for text style transfer. The first one is the backpropagation over the discrete sample. We can use gumbel-softmax distribution as a continuous approximation. The other one is the lack of evaluation benchmark, such as style accuracy, fluency, and embedding similarity.

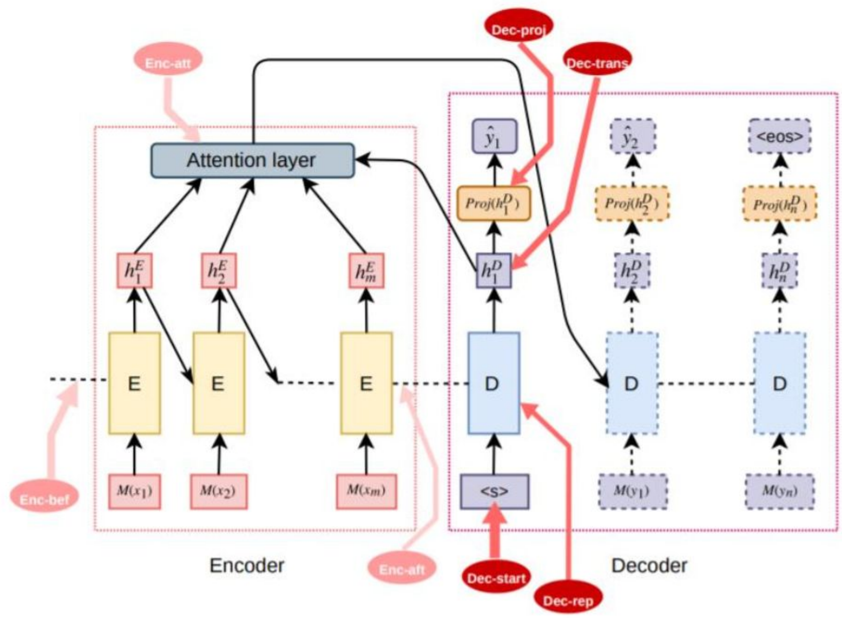
- **Emotional Conversation Overview**

  It is a difficult task to generate a sentence with target emotion. Besides **ECM,** there are still other works. Considering the topic and emotion in a reply, we can generate a sentence with topic first and emotion last or vice versa. Therefore, we first extract two words about topic and emotion which will appear in the replies. After that, use the attention hidden state of these two words to generate middle sentence. Finally, concatenate emotion keywords, topic keywords, and middle sentence to produce whole replies. This method is a novel and new approach. (Fig.6)



$$y = y^t + w_{tp} + y^{md} + w_{et} + y^e$$

Another work is to try different concatenation of emotion embedding on model architecture, such as encoder first, decoder first, decoder each time step input, decoder each time step output, decoder each time step before selection of words, and decoder attention with encoder.



- **Emotional Conversation Implementation**

In my first attempt, I simply use embedding concatenation on decoder to extract different emotion sentences. However, my model would always give some common and repeated words, which cannot show a good result.

```
Your Sentence:你好阿
Label (0: Other, 1: Like, 2: Sadness, 3: Disgust, 4: Anger, 5: Happiness):0
> BOS你好阿PADPADPADPADPADPADPADPADPADPADPADPADPADPADPADPADPADPADPADPADPADPADPADPADPADPADEOS
< 幸好幸好吶吶吶吶吶吶吶吶吶吶吶吶吶吶吶吶吶吶吶吶吶吶吶吶吶吶吶吶
Your Sentence:你好阿
Label (0: Other, 1: Like, 2: Sadness, 3: Disgust, 4: Anger, 5: Happiness):1
> BOS你好阿PADPADPADPADPADPADPADPADPADPADPADPADPADPADPADPADPADPADPADPADPADPADPADPADPADPADEOS
< 幸好吶吶吶吶吶吶吶吶吶吶吶吶吶吶吶吶吶吶吶吶吶吶吶吶吶吶吶吶吶
Your Sentence:你好阿
Label (0: Other, 1: Like, 2: Sadness, 3: Disgust, 4: Anger, 5: Happiness):2
> BOS你好阿PADPADPADPADPADPADPADPADPADPADPADPADPADPADPADPADPADPADPADPADPADPADPADPADPADPADEOS
< 果断本事本事本事本事本事本事本事本事本事本事本事本事本事本事本事本事本事本事本事本事本事本事
事本事本事
Your Sentence:你好阿
Label (0: Other, 1: Like, 2: Sadness, 3: Disgust, 4: Anger, 5: Happiness):3
> BOS你好阿PADPADPADPADPADPADPADPADPADPADPADPADPADPADPADPADPADPADPADPADPADPADPADPADPADPADEOS
< 果断本事本事本事本事本事本事本事本事本事吶吶吶吶吶本事吶本事吶吶吶吶吶本事吶本事吶本事吶本事
Your Sentence:你好阿
Label (0: Other, 1: Like, 2: Sadness, 3: Disgust, 4: Anger, 5: Happiness):4
> BOS你好阿PADPADPADPADPADPADPADPADPADPADPADPADPADPADPADPADPADPADPADPADPADPADPADPADPADPADEOS
< 果断吶吶吶吶吶吶吶吶吶吶吶吶吶吶吶吶吶吶吶吶吶吶吶吶吶吶吶吶吶
Your Sentence:你好阿
Label (0: Other, 1: Like, 2: Sadness, 3: Disgust, 4: Anger, 5: Happiness):5
> BOS你好阿PADPADPADPADPADPADPADPADPADPADPADPADPADPADPADPADPADPADPADPADPADPADPADPADPADPADEOS
< 幸好幸好吶吶吶吶吶吶吶吶吶吶吶吶吶吶吶吶吶吶吶吶吶吶吶吶吶吶吶吶
```

In my second attempt, I use beam search to extract more different sentences, which can give a longer one. However, it still cannot show the difference of emotions.

Therefore, I use ECM to verify their work in my last experiment. The results can give diversities on emotions. However, it is still a challenge to get rid of common sentences.

Post	Emotion	Response
發大財!	None	有 是 可以 做 的 。
	Like	好 的 啊 ， 我 的
	Sadness	唉 ， 孩子 啊 ！
	Disgust	你 的 的 厚道 瞭 ！
	Anger	你 的 的 ！ ！ ！
	Happiness	哈哈 ， ~ 啊 ~
Post	**Emotion**	**Response**
廢物	None	說 說 不 不 知道
	Like	很 好 的 ！ ！
	Sadness	可憐 很 ， 啊 ！
	Disgust	你 啊 不 厚道 瞭 ！
	Anger	你 的 ！ ！ ！
	Happiness	哈哈 ！ ！ ！ ~

❑ Future Works

I think the most difficulties for my research is that it is hard to find some resources such as code or works in emotion conversation. Also, whether to build a chatbot or a style transfer model is hard to define. I think I have to pay my whole time on this research trying some methods and finding some novelty.

Perhaps the next step is to understand the ECM reference code and use the transformer as the seq2seq model. If it still cannot give good results on chatbot, I will try style transfer on emotion rather than end to end training on emotion chatbot.

❑ Reference

[1] Fu, Zhenxin, et al. "Style transfer in text: Exploration and evaluation." *Thirty-Second AAAI Conference on Artificial Intelligence*. 2018.

[2] Shen, Tianxiao, et al. "Style transfer from non-parallel text by cross-alignment." *Advances in neural information processing systems*. 2017.

[3] Hu, Zhiting, et al. "Toward controlled generation of text." *Proceedings of the 34th International Conference on Machine Learning-Volume 70*. JMLR. org, 2017.

[4] John, Vineet, et al. "Disentangled representation learning for text style transfer." *arXiv preprint arXiv:1808.04339* (2018).

[5] Li, Jingyuan, and Xiao Sun. "A Syntactically Constrained Bidirectional-Asynchronous Approach for Emotional Conversation Generation." *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing*. 2018.

[6]Huang, Chenyang, et al. "Automatic dialogue generation with expressed emotions." *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 2 (Short Papers)*. 2018.