自然语言生成关键技术研究

自然语言生成

- 逻辑上输出为文本的任务均需要自然语言生成技术
- 如**文本到文本**的生成、**图像到文本**的生成

自动文摘：
根据输入文档自动生成摘要

昨晚，中联航空成都飞北京一架航班被发现有多人吸烟。后因天气原因，飞机备降太原机场。有乘客要求重新安检，机长决定继续飞行，引起机组人员与未吸烟乘客冲突。

⇩

航班多人吸烟机组人员与乘客 冲突

数据来源：LCSTS [Hu et al., 2015]

图像描述生成:
根据输入图像生成描述性文字

⇩

A man stands by a rainy street with an umbrella over his head

数据来源：COCO [Lin et al., 2014]

研究领域

- **研究内容十分广泛、任务十分多样**
- **文本至文本**
  - 机器翻译[Hutchins & Somers 1992]、自动文摘[Clarke & Lapata, 2010]、文本简化[Siddharthan, 2014]、自动改错[Kukich, 1992]、学术论文评审意见生成[Bartoli et al., 2016]、文本复述[Bannard & Callison-Burch, 2005]、问题生成[Brown et al., 2005]、风格转换
- **数据至文本**
  - 体育比赛报告生成[Theune et al., 2001]、新闻速览[Lepp et al., 2017]、天气预报[Goldberg et al. , 1994]、博物馆导览[O'Donnell, 2001]等
- **视觉至文本**
  - 图片描述生成[Farhadi et al., 2010]、视觉问答[Antol et al., 2015]

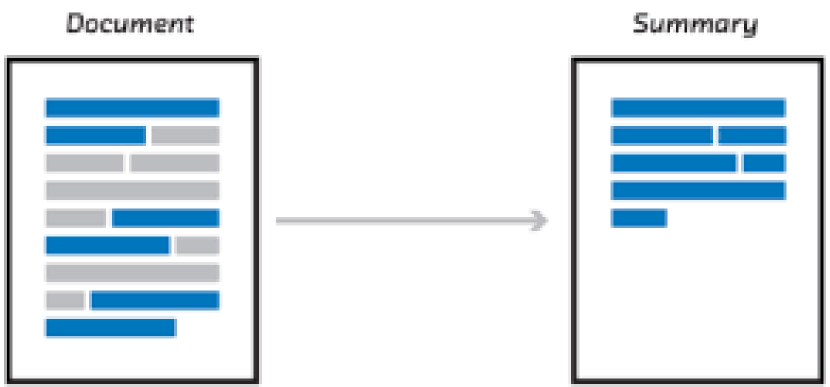
研究领域 – 文本至文本生成

- **文本至文本生成研究更为充分，很多都成为了经典的、单独的、有影响力的研究领域**
  - 比如机器翻译、自动文摘等
  - 这也是为什么NLG定义一般不包括这些较为独立的任务

- **机器翻译 (Machine Translation, MT)**
  - 将一种语言用另一种语言表示
  - 研究自计算机开始即开始
  - 有影响力的评测包括NIST、WMT等
  - 有影响力的模型包括
    - 噪音信道模型
    - IBM的Model I, II, III
    - 序列到序列模型

研究领域 – 文本至文本生成

- **自动文摘 (automatic text summarization, AS)**
  - 将文档用几句话总结
  - 研究始于NIST组织的DUC系列会议、后成为TAC
    - 面向单文档、多文档的摘要
      - 多文档摘要的核心难点是实体聚类、消歧；冲突化解
    - 面向焦点对象的摘要

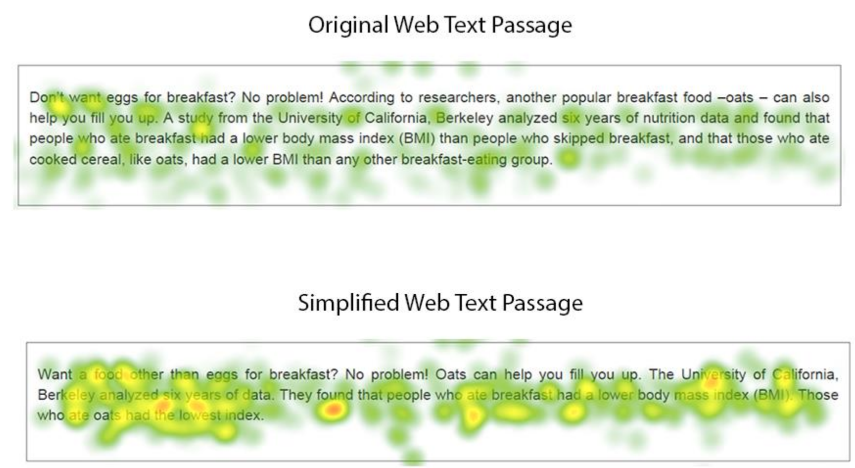


  - 现今研究倾向于将AS划分为
    - 抽取式摘要
    - 生成式摘要
  - 常用数据集有Giga, CNN, LCSTS等
    - 胜在数量、但质量远逊于DUC

研究领域 – 文本至文本生成

文本简化 (automatic text simplification)

- 让文本更通俗易懂
- 服务儿童或有阅读障碍群体
- 提高网页的停留时长
  - eye tracking显示人的关注点在那些简单、常见的词上
- 常用数据集有
  - PKWP
  - EW-SEW
  - 均是抽取的Wikipedia和Simple Wikipedia的对齐语料



研究领域 – 多模态语言生成

- **最近几年，图像与语言结合的任务迅速兴起**
  - CV与NLP结合的交叉领域
  - 寻找语言的感知基础一直是AI的科学关切
- **目前有两个热点: Image captioning和Visual QA**
- **图像标题生成 (Image Captioning)**
  - 给定一个图片生成合适的**描述性**文字
    - “这是张猫的图片”，“白猫的例子”这样的标题不作数
  - 有影响力的评测
    - MSCOCO captioning track
  - 常用数据集还有Flickr30k

a man stands by a rainy street with an umbrella over his head.
a middle aged lady standing by the curb and holding an umbrella.
a man in yellow and shorts is on the side of the road with an umbrella.
street intersection with stopped cars and a pedestrian holding an umbrella on a rainy day.
a person with an umbrella begins to cross the rain covered road.

研究领域 –多模态语言生成

- 图像问答(Visual Question Answering)
  - 根据图像回答问题
  - 探究语言与感知信息的关系
  - 有影响力的评测有VirginiaTech和GeorgiaTech组织的VQA Challenge

Question : What color is the hydrant?

Original Image | red

Complementary Image | black and yellow

核心问题

- 共同的核心问题：说什么与如何说

- 说什么
  - 输入中的哪些内容需要在输出中体现

- 如何说
  - 如何将选定的内容组织为自然语言

电影评论摘要

原文

I saw this movie 11 times in the theater and I think that this is one of the best movies ever made and the best movie made about Christ and his passion. God bless all those responsible for the creation of this powerful film.

⇩

重要性

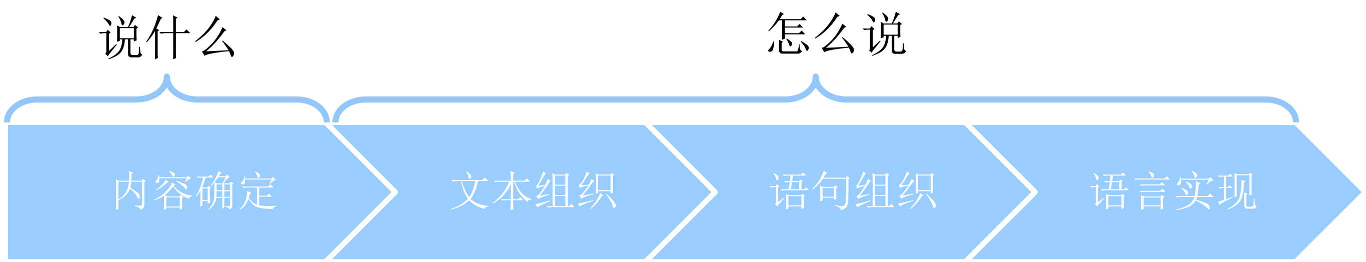
I saw this movie 11 times in the theater and I think that this is one of the best movies ever made and the best movie made about Christ and his passion. God bless all those responsible for the creation of this powerful film.

⇩

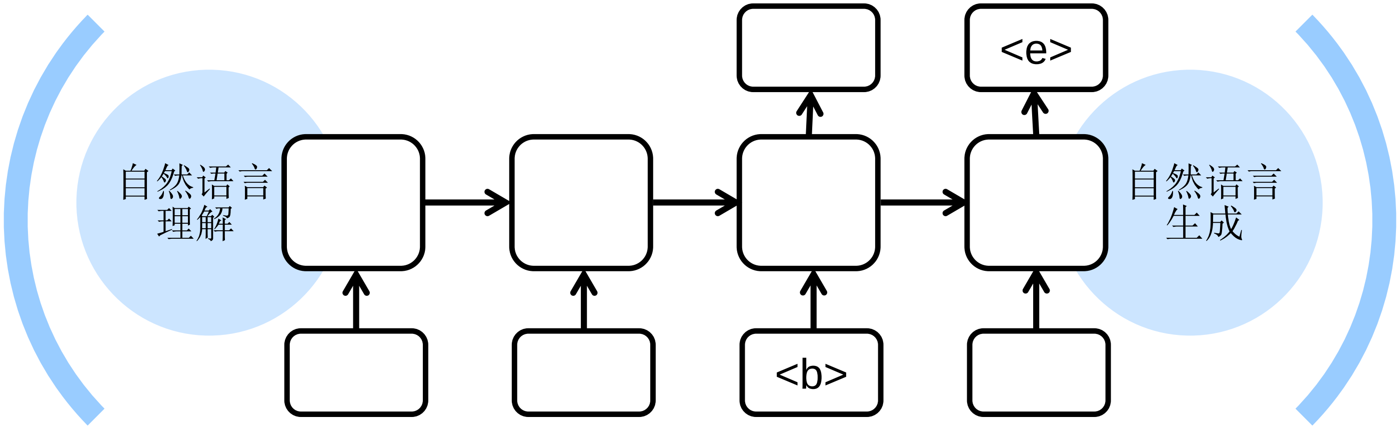
摘要 Best movie about Christ

发展进程

- 传统方法：**分步**建模的复杂系统



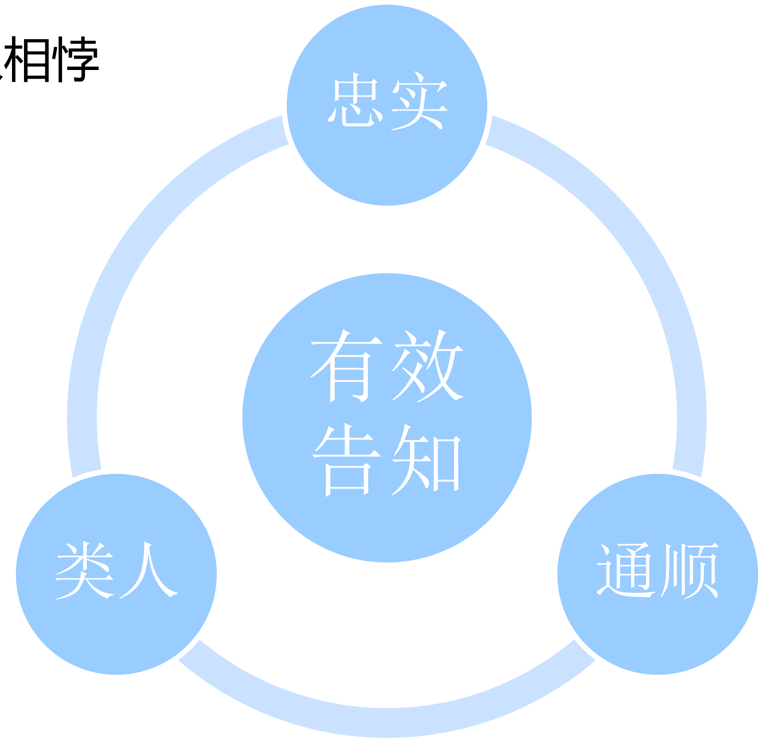
- 深度学习方法：简单直观的**端到端**模型



主要挑战

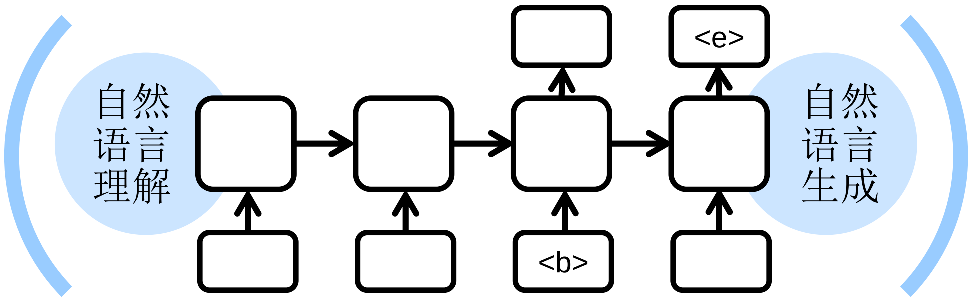
- **自然语言生成的主要目的是有效的告知：说得好**

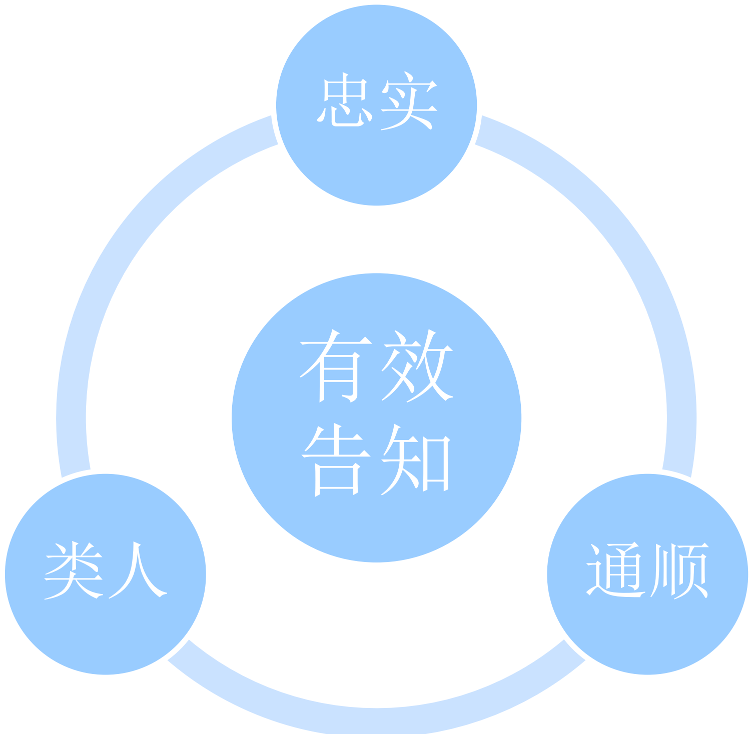
- 内容应当准确忠实
  - **一致性**：不应语句前后矛盾、或与输入相悖
- 语言应通顺易懂
  - **流畅性**：不应不成句子
- 尽可能符合人类表达习惯
  - **多样性**：表达不应过于单一
  - **感情色彩**：尽可能提现情感倾向



主要挑战

- **深度学习方法**
  - 极大的简化了模型设计，特别是神经语言模型基本解决了通顺度问题
  - 大幅提升了机器翻译、自动文摘等任务的效果
  - 但仍未完全达到人类的语言的效果

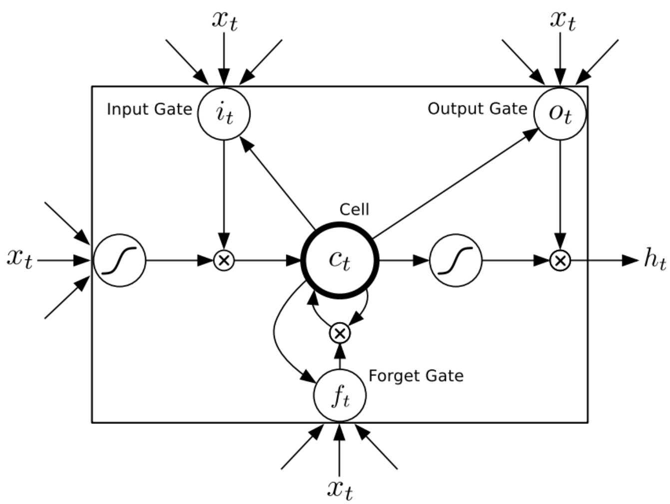




- **特别是端到端模型模糊了说什么和如何说的界限**
  - 使得有针对性的改进更为困难

神经语言模型

- Graves (2013)展示了**word-level LSTM**在语言建模中的效果



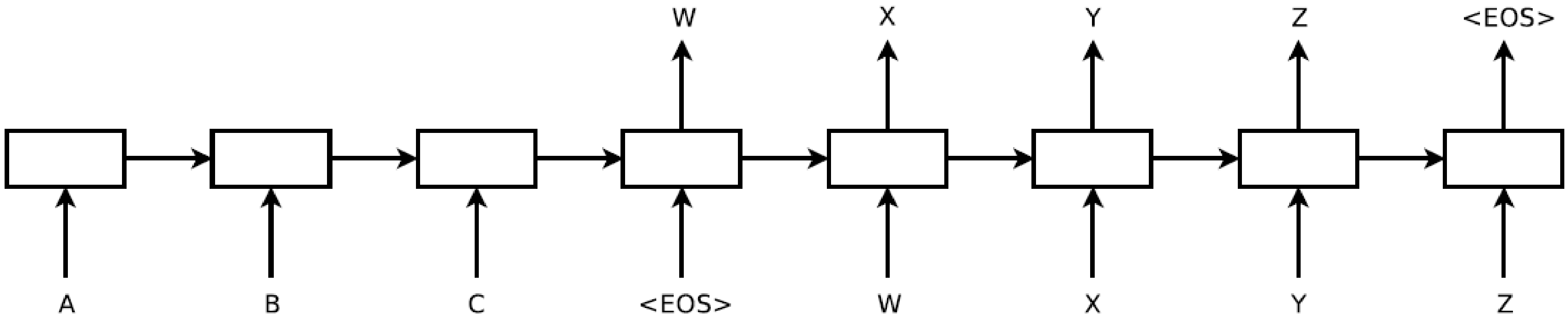
Method	Bits per Character
bzip2	2.32
M-RNN[1]	1.6 (text only)
deep LSTM	**1.42** (1.33 validation)
PAQ-8[2]	1.28

  - Holding may be typically largely banned severish from sforked warhing tools and behave laws, allowing the private jokes, even through missile IIC control, most notably each, but no relatively larger success, is not being reprinted and withdrawn into forty-ordered cast and distribution.

- 但如何将这些free-run方法用于自然语言生成?

编码器解码器框架

- **更进一步，将语言理解过程融入模型**
  - 一个彻底的**端到端**模型
- **编码器解码器框架最早由Sutskever et al. (2014)提出**
  - Encoder RNN将输入编码为向量表示，作为Decoder RNN的额外输入
  - 这一架构非常适合**序列到序列**(Sequence-to-Sequence, Seq2Seq)类型的任务，例如机器翻译[Kalchbrenner & Blunsom, 2013; Bahdanau et al., 2015]



编码器解码器框架

- 更进一步，出现了基于**注意力机制**的模型[Bahdanau et al. 2015; Luong et al., 2015]
  - 注意力学习到了输入表示到输出文本的松散耦合(loose couplings)

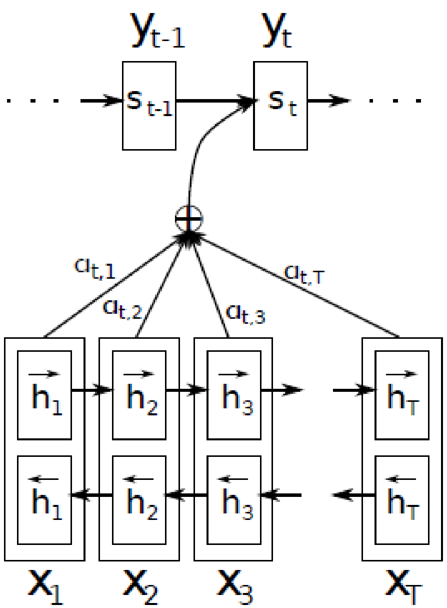
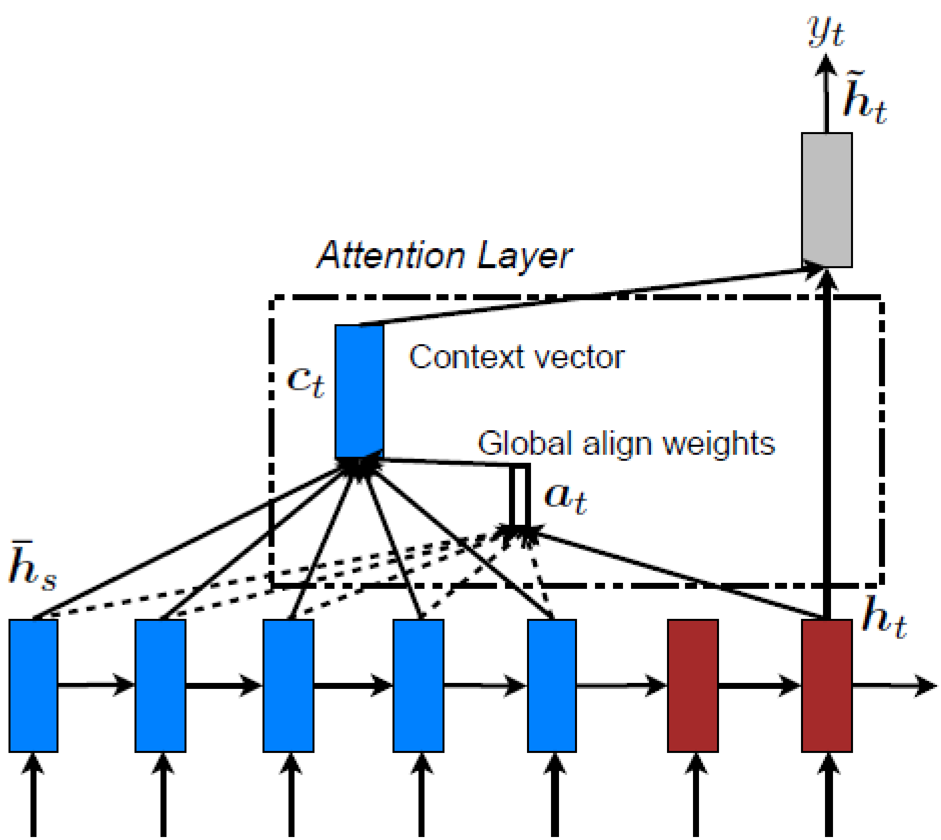
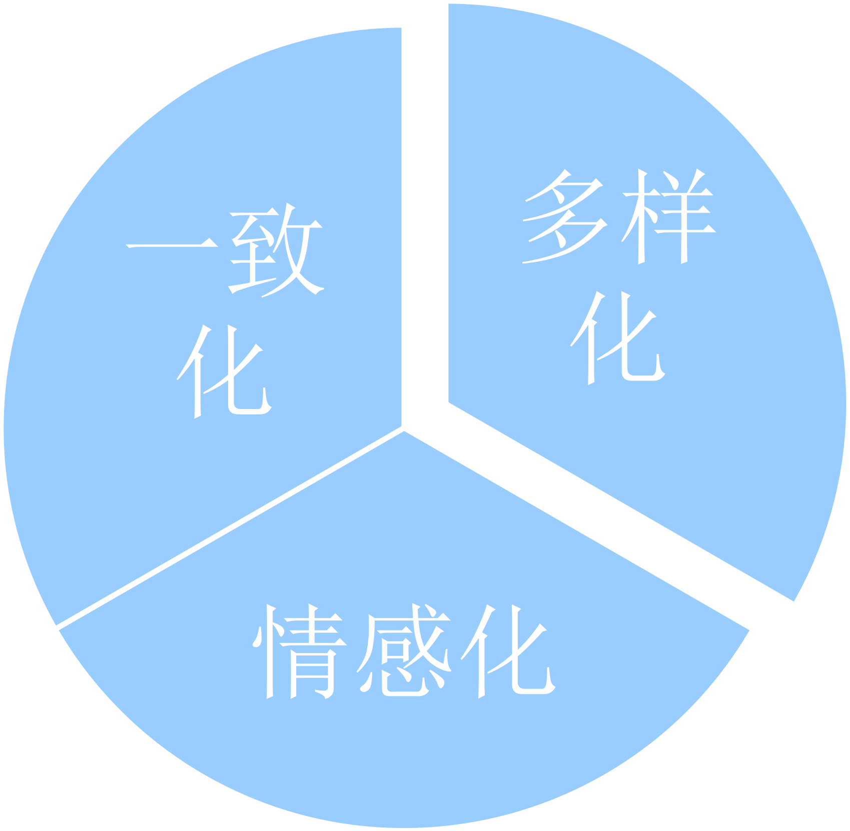


Figure 1: The graphical illustration of the proposed model trying to generate the t-th target word y_t given a source sentence $(x_1, x_2, \ldots, x_T)$.



主要挑战

- 让模型说得好



多样化自然语言生成

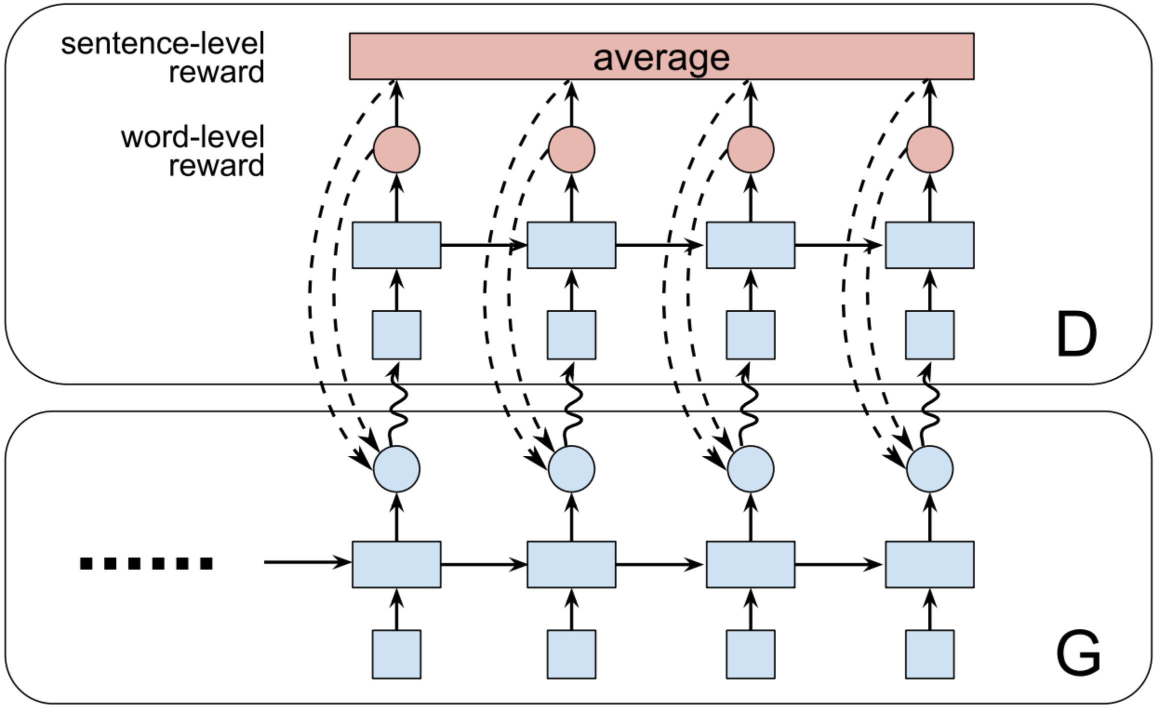
- **问题：**
  - 最大似然方法训练导致模型倾向生成数据中高频表达
  - 生成文本**表达单一**，缺少信息量
    - 对话中，回复“I am sorry”

词	真实文本频率	生成文本频率
the	4.2%	7.1%
and	3.2%	4.6%
was	1.5%	5.3%

- **如何改进模型输出多样性？**
  - 语言模型可以衡量一个句子的概率
  - P(多样性表达句子)<P(单一表达句子)
- **通过强化学习方法鼓励模型生成低概率的句子**
  - $R(y_t) = -\frac{1}{K}\sum_{k=1}^{K} \log P\left(y_{t,k} \middle| y_{t,<k}\right)$ -- Reward
  - 但低概率的句子也可能是那些真正错误的句子

多样化自然语言生成

- DP-GAN (Diversity-Promoting GAN) [Xu et al, 2018]
- 利用对抗学习的思想
  - 生成器目标：生成多样化、真实的句子
  - 判别器目标：给多样化、真实的句子更高的分数
    - 判别器由语言模型充当



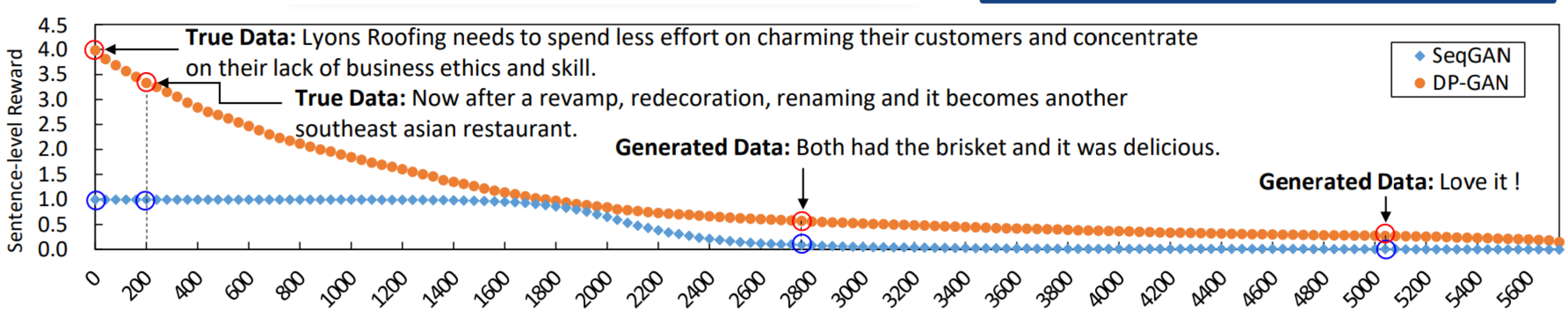
对抗强化学习

- 困惑度作为奖励函数
  - 句子级别困惑度
    - $\mathrm{R}(y_t) = -\frac{1}{K}\sum_{k=1}^{K} logD_{\emptyset}(y_{t,k}|\boldsymbol{y}_{<t,k})$
  - 单词级别困惑度
    - $\mathrm{R}(y_{t,k}|\boldsymbol{y}_{<t,k}) = -logD_{\emptyset}(y_{t,k}|\boldsymbol{y}_{<t,k})$
- 优点
  - 奖励函数更加平滑
  - 缓解奖励饱和问题

Policy Gradient Training

Adversarial reinforcement training:

Initialize G_θ, D_ϕ with random weights θ, ϕ
Pre-train G_θ using MLE on a sequence dataset $\mathcal{D} = (X, Y)$
Generate samples using G_θ for training D_ϕ
Pre-train D_ϕ
N = number of training iterations
M = number of training generator
K = number of training discriminator
for each $i = 1, 2, ..., N$ **do**
  for each $j = 1, 2, ..., M$ **do**
    Generate a sequence $Y_{1:T} \sim G_\theta$
    Update generator via policy gradient
    Sample a sequence $Y_{1:T} \sim \mathcal{D}$
    Update generator parameters
  end for
  for each $j = 1, 2, ..., K$ **do**
    Generate samples using G_θ
    Train discriminator D_ϕ
  end for
end for



多样化自然语言生成

- DP-GAN (Diversity-Promoting GAN) [Xu et al, 2018]
- 实验数据
  - Yelp Review Generation
  - Amazon Review Generation
  - OpenSubtitles Dialogue
- 自动评价
  - 生成文本更长
  - 生成表达更丰富（独特的ngram数量）
- 人工评价
  - DP-GAN文本排名更高

Yelp	Token	Dist-1	Dist-2	Dist-3	Dist-S
MLE	151.2K	1.2K	3.9K	6.6K	3.9K
PG-BLEU	131.1K	1.1K	3.3K	5.5K	3.1K
SeqGAN	140.5K	1.1K	3.5K	6.1K	3.6K
DP-GAN(S)	**438.6K**	1.7K	7.5K	15.7K	10.6K
DP-GAN(W)	271.9K	2.8K	14.8K	29.0K	12.6K
DP-GAN(SW)	406.8K	**3.4K**	**22.3K**	**49.6K**	**17.3K**
Amazon	Token	Dist-1	Dist-2	Dist-3	Dist-S
MLE	176.1K	0.6K	2.1K	3.5K	2.6K
PG-BLEU	124.5K	0.6K	1.9K	3.5K	2.3K
SeqGAN	217.3K	0.7K	2.6K	4.6K	3.2K
DP-GAN(S)	**467.6K**	0.8K	3.6K	7.6K	7.0K
DP-GAN(W)	279.4K	1.6K	8.9K	18.4K	9.6K
DP-GAN(SW)	383.6K	**1.9K**	**11.7K**	**26.3K**	**13.6K**
Dialogue	Token	Dist-1	Dist-2	Dist-3	Dist-S
MLE	81.1K	1.4K	4.4K	6.3K	4.1K
PG-BLEU	97.9K	1.2K	3.9K	5.5K	3.3K
SeqGAN	83.4K	1.4K	4.5K	6.5K	4.5K
DP-GAN(S)	**112.2K**	1.5K	5.2K	8.5K	5.6K
DP-GAN(W)	79.4K	1.9K	7.7K	11.4K	6.0K
DP-GAN(SW)	97.3K	**2.1K**	**10.8K**	**19.1K**	**8.0K**

	Model	Averaged Ranking
Yelp	MLE	1.89
	PG-BLEU	2.22
	SeqGAN	2.12
	DP-GAN	**1.51**
Amazon	MLE	1.93
	PG-BLEU	2.24
	SeqGAN	1.98
	DP-GAN	**1.50**
Dialogue	MLE	2.46
	PG-BLEU	2.40
	SeqGAN	2.17
	DP-GAN	**1.92**

多样化自然语言生成

- DP-GAN (Diversity-Promoting GAN) [Xu et al, 2018]
- 实验数据
  - Yelp Review Generation
  - Amazon Review Generation
  - OpenSubtitles Dialogue
- 自动评价
  - 生成文本更长
  - 生成表达更丰富（独特的ngram数量）
- 人工评价
  - DP-GAN文本排名更高

Yelp	Token	Dist-1	Dist-2	Dist-3	Dist-S
MLE	151.2K	1.2K	3.9K	6.6K	3.9K
PG-BLEU	131.1K	1.1K	3.3K	5.5K	3.1K
SeqGAN	140.5K	1.1K	3.5K	6.1K	3.6K
DP-GAN(S)	**438.6K**	1.7K	7.5K	15.7K	10.6K
DP-GAN(W)	271.9K	2.8K	14.8K	29.0K	12.6K
DP-GAN(SW)	406.8K	**3.4K**	**22.3K**	**49.6K**	**17.3K**
Amazon	Token	Dist-1	Dist-2	Dist-3	Dist-S
MLE	176.1K	0.6K	2.1K	3.5K	2.6K
PG-BLEU	124.5K	0.6K	1.9K	3.5K	2.3K
SeqGAN	217.3K	0.7K	2.6K	4.6K	3.2K
DP-GAN(S)	**467.6K**	0.8K	3.6K	7.6K	7.0K
DP-GAN(W)	279.4K	1.6K	8.9K	18.4K	9.6K
DP-GAN(SW)	383.6K	**1.9K**	**11.7K**	**26.3K**	**13.6K**
Dialogue	Token	Dist-1	Dist-2	Dist-3	Dist-S
MLE	81.1K	1.4K	4.4K	6.3K	4.1K
PG-BLEU	97.9K	1.2K	3.9K	5.5K	3.3K
SeqGAN	83.4K	1.4K	4.5K	6.5K	4.5K
DP-GAN(S)	**112.2K**	1.5K	5.2K	8.5K	5.6K
DP-GAN(W)	79.4K	1.9K	7.7K	11.4K	6.0K
DP-GAN(SW)	97.3K	**2.1K**	**10.8K**	**19.1K**	**8.0K**

	Model	Averaged Ranking
Yelp	MLE	1.89
	PG-BLEU	2.22
	SeqGAN	2.12
	DP-GAN	**1.51**
Amazon	MLE	1.93
	PG-BLEU	2.24
	SeqGAN	1.98
	DP-GAN	**1.50**
Dialogue	MLE	2.46
	PG-BLEU	2.40
	SeqGAN	2.17
	DP-GAN	**1.92**

多样化自然语言生成

- DP-GAN (Diversity-Promoting GAN) [Xu et al, 2018]
- 实验数据
  - Yelp Review Generation
  - Amazon Review Generation
  - OpenSubtitles Dialogue
- 自动评价
  - 生成文本更长
  - 生成表达更丰富（独特的ngram数量）
- 人工评价
  - 对比排序
  - DP-GAN文本排名更高

Yelp	Token	Dist-1	Dist-2	Dist-3	Dist-S
MLE	151.2K	1.2K	3.9K	6.6K	3.9K
PG-BLEU	131.1K	1.1K	3.3K	5.5K	3.1K
SeqGAN	140.5K	1.1K	3.5K	6.1K	3.6K
DP-GAN(S)	**438.6K**	1.7K	7.5K	15.7K	10.6K
DP-GAN(W)	271.9K	2.8K	14.8K	29.0K	12.6K
DP-GAN(SW)	406.8K	**3.4K**	**22.3K**	**49.6K**	**17.3K**
Amazon	Token	Dist-1	Dist-2	Dist-3	Dist-S
MLE	176.1K	0.6K	2.1K	3.5K	2.6K
PG-BLEU	124.5K	0.6K	1.9K	3.5K	2.3K
SeqGAN	217.3K	0.7K	2.6K	4.6K	3.2K
DP-GAN(S)	**467.6K**	0.8K	3.6K	7.6K	7.0K
DP-GAN(W)	279.4K	1.6K	8.9K	18.4K	9.6K
DP-GAN(SW)	383.6K	**1.9K**	**11.7K**	**26.3K**	**13.6K**
Dialogue	Token	Dist-1	Dist-2	Dist-3	Dist-S
MLE	81.1K	1.4K	4.4K	6.3K	4.1K
PG-BLEU	97.9K	1.2K	3.9K	5.5K	3.3K
SeqGAN	83.4K	1.4K	4.5K	6.5K	4.5K
DP-GAN(S)	**112.2K**	1.5K	5.2K	8.5K	5.6K
DP-GAN(W)	79.4K	1.9K	7.7K	11.4K	6.0K
DP-GAN(SW)	97.3K	**2.1K**	**10.8K**	**19.1K**	**8.0K**

	Model	Averaged Ranking
Yelp	MLE	1.89
	PG-BLEU	2.22
	SeqGAN	2.12
	DP-GAN	**1.51**
Amazon	MLE	1.93
	PG-BLEU	2.24
	SeqGAN	1.98
	DP-GAN	**1.50**
Dialogue	MLE	2.46
	PG-BLEU	2.40
	SeqGAN	2.17
	DP-GAN	**1.92**

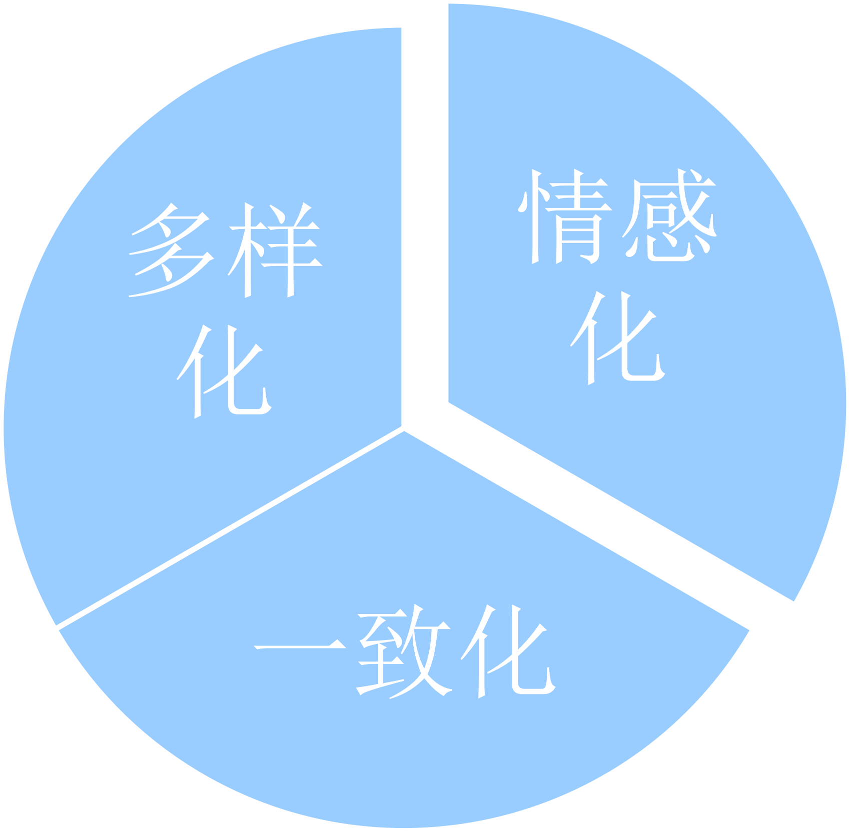
- **DP-GAN (Diversity-Promoting GAN) [Xu et al, 2018]**
- **生成样例**

Input: *One of my favorite places to eat.*
MLE: *Service is great.*
PG-BLEU: *Service is always good.*
SeqGAN: *Love the chicken and waffles. Service is always great.*
DP-GAN: *Love the fact that they have a large selection of food. Service is always great and the food is always fresh. I've been to this place a few times and have never been disappointed.*

Input: *I don't think it's too spicy, so i add a little hot curry spice to the sauce.*
MLE: *It's great for cooking. I have to say, I'm not sure if it is the same.*
PG-BLEU: *Love it!*
SeqGAN: *Love it! Love this stuff. I have used it for years and it is very good.*
DP-GAN:*I've tried many different brands of hot sauces but this one is the best. I've also used it in soups and stews. I also like that it is organic and has a great flavor.*

主要挑战

- 让模型说得好



情感化自然语言生成

- **样例:**

1) **The movie is amazing! — The movie is boring!**

2) **I went to this restaurant last weak, the staff was friendly, and I were so happy to have a great meal! — I went to this restaurant last weak, the staff was rude, and I were so angry to have a terrible meal!**

定义

情感转换的目标是在不改变句子内容语义的情况下，改变句子情感。

情感化自然语言生成

应用：对话机器人

I am sad about the failure of the badminton player A.

The badminton player B defeats A. Congratulations!

sentiment-to-sentiment translation

Refined Answer: I’m sorry to see that the badminton player B defeats A.

情感化自然语言生成

应用：个性化新闻写作

Sentiment-to-sentiment translation can save a lot of human labor!

The visiting team defeated the home team

News for fans of the visiting team: The players of the home team performed badly, and lost this game.

News for fans of the home team: Although the players of the home team have tried their best, they lost this game regretfully.

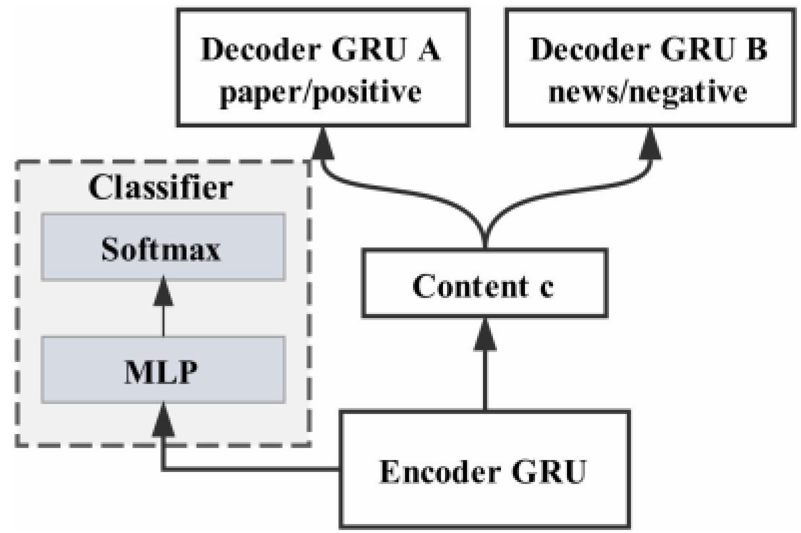
情感化自然语言生成

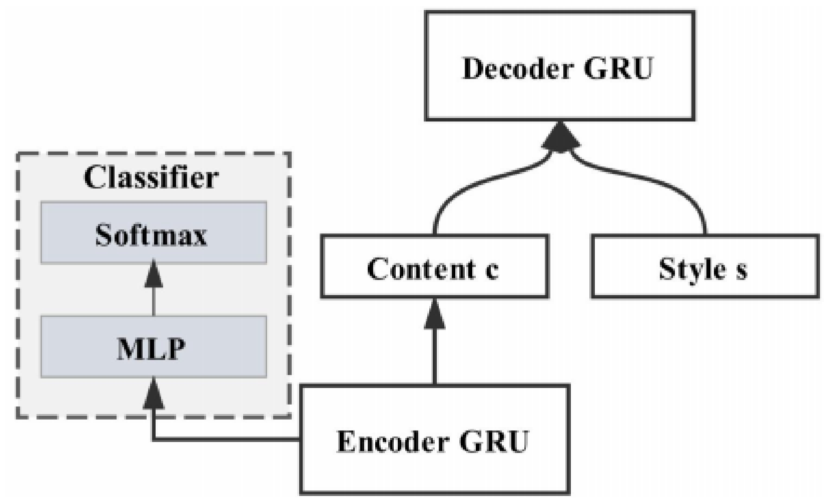
- **情感化语言生成的一个重要问题是如何转换句子的情感极性**
- **主要挑战：缺少有监督的数据（平行语料）**
  - 但有大量的非平行语料：购物、点评网站上的文本及其对应的评分
  - 如果模型能够根据这类非平行语料转换文本情感，那模型应具备了情感化文本生成的能力
- **要完成这项任务**
  - 模型需要了解语言的**情感构成**
    - 哪些是情感相关内容
    - 哪些是非情感内容
  - 需要了解文本**情感极性的影响因素**
    - 哪些因素是积极倾向的
    - 哪些因素是消极倾向的

句子情感极性转换是情感建模能力的试金石

情感化自然语言生成

- **如何让模型更好的区分情感内容和非情感内容？**
  - 现有方法期望编码器编码后的内容有较独立的内容信息和情感信息
  - 但现有编码器解码器模型难以做到这一点
  - 修改情感信息会牵连到内容信息





图稿来源：[Fu et al., 2018]

情感化自然语言生成

现有工作

- 现有模型能够**较好地转换极性**
- 但转换后内容往往**与输入无关**
- 模型没有理解文本的情感构成

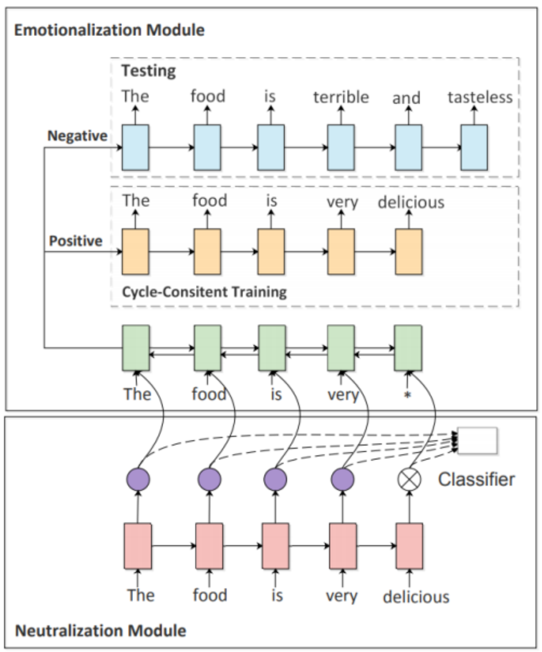
The food is delicious

→ What a bad movie

图稿来源：[Fu et al., 2018]

情感转换 Sentiment conversion

- 提出**弱监督强化学习模型**，提高了情感转换效果



Input: *I would strongly advise against using this company.*
CAAE: *I love this place for a great experience here.*
MDAL: *I have been a great place was great.*
Proposed Method: *I would love using this company.*

Input: *The service was nearly non-existent and extremely rude.*
CAAE: *The best place in the best area in vegas.*
MDAL: *The food is very friendly and very good.*
Proposed Method: *The service was served and completely fresh.*

Input: *Asked for the roast beef and mushroom sub, only received roast beef.*
CAAE: *We had a great experience with.*
MDAL: *This place for a great place for a great food and best.*
Proposed Method: *Thanks for the beef and spring bbq.*

Input: *Worst cleaning job ever!*
CAAE: *Great food and great service!*
MDAL: *Great food, food!*
Proposed Method: *Excellent outstanding job ever!*

挑战：词典是否可以解决？

- 简单的反义词替换导致句子不流畅

The food is terrible like rock

The food is delicious like rock

挑战：词典是否可以解决？

- **很多单词是有歧义的，需要额外的词语消歧**
  - 比如, “good”在WordNet有三个反义词: “evil”, “bad”, and “ill”. 选择哪个词作为替换需要充分对上下文进行理解。

evil

ill

bad

情感话自然语言生成

◻ 方法介绍

◻ 去情感模块

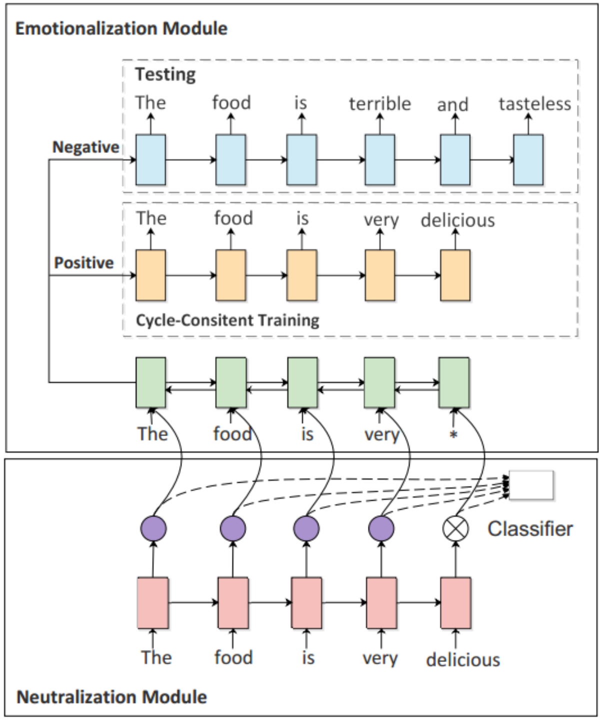
- ➢使用attention机制mask情感词语
- ➢情感分类和序列标注联合训练

◻ 加情感模块

- ➢情感化语言生成模型
- ➢条件生成器

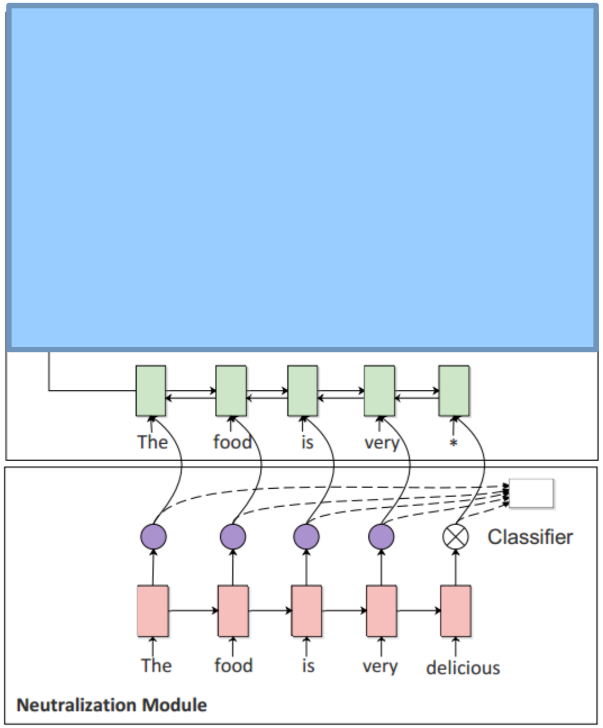
◻循环一致训练

- ➢循环强化一致训练



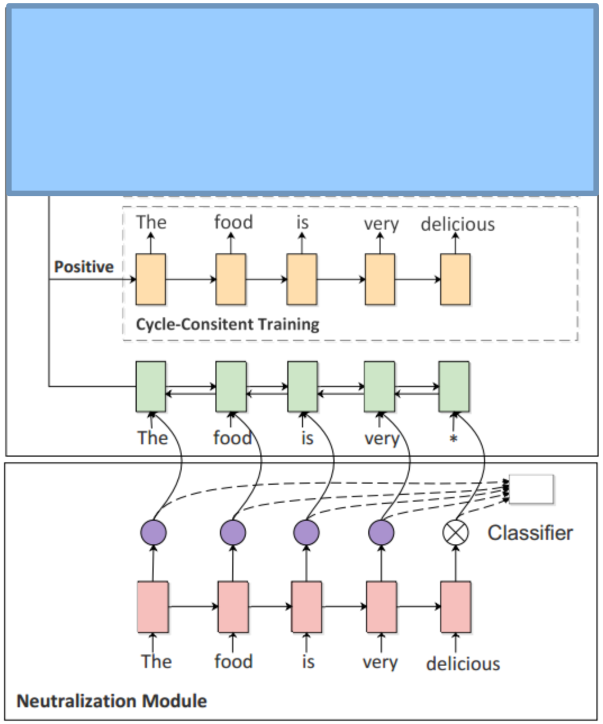
循环强化一致训练

- 1) 将情感化的句子中性化
- 2) 增加同样的情感重构输入的情感话句子.
- 3) 使用重建误差训练情感化模型
- 4) 使用强化学习训练去情感化模块



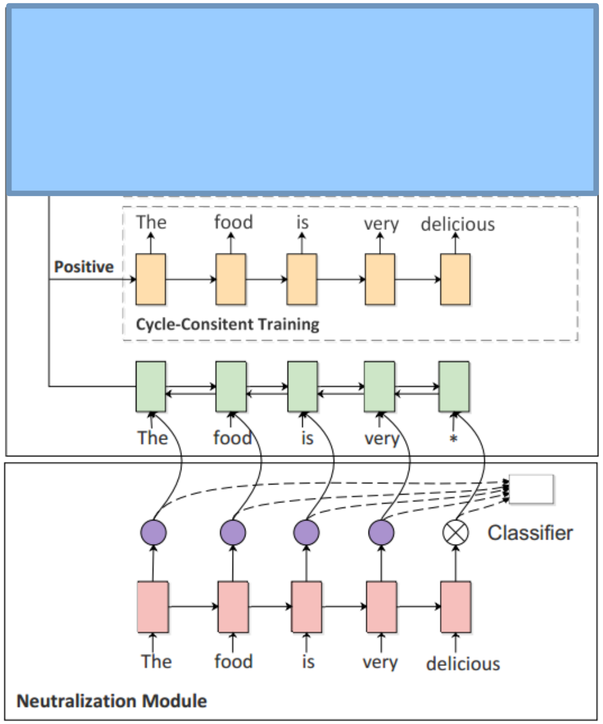
循环强化一致训练

- 1) 将情感化的句子中性化
- 2) 增加同样的情感重构输入的情感化句子
- 3) 使用重建误差训练情感化模型
- 4) 使用强化学习训练去情感化模块



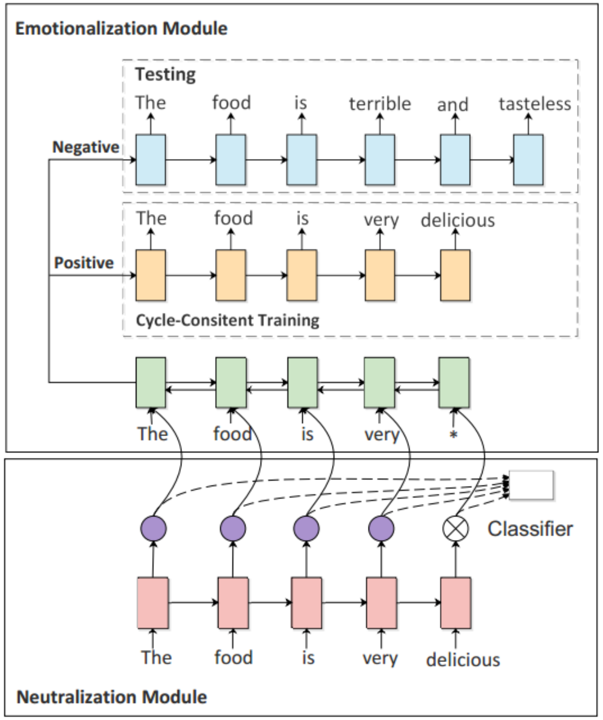
循环强化一致训练

- 1) 将情感化的句子中性化
- 2) 增加同样的情感重构输入的情感化句子
- **3) 使用重建误差训练情感化模型**
- 4) 使用强化学习训练去情感化模块



循环强化一致训练

- 1) 将情感化的句子中性化
- 2) 增加同样的情感重构输入的情感化句子
- 3) 使用重建误差训练情感化模型
- **4) 使用强化学习训练去情感化模块**



数据集

Yelp Review Dataset (Yelp)

- Yelp Dataset Challenge

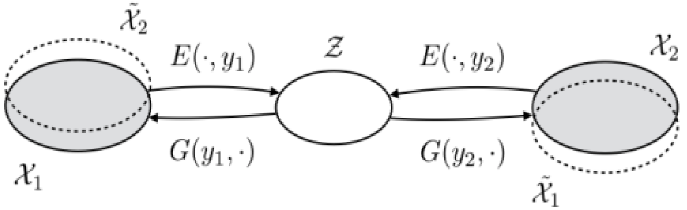
Amazon Food Review Dataset (Amazon)

- Provided by McAuley and Leskovec (2013). It consists of amounts of food reviews from Amazon

比较方法

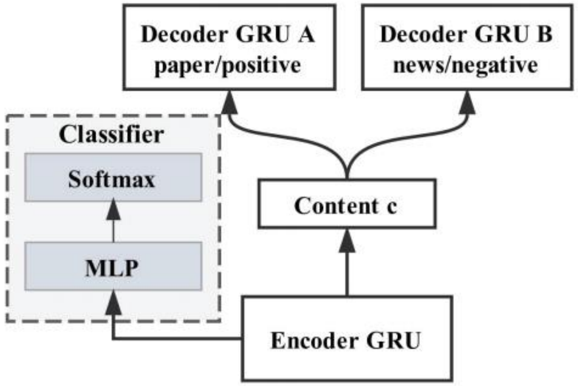
Cross-Alignment Auto-Encoder (CAAE)

- Refined alignment of latent



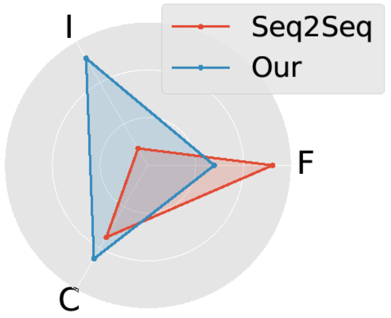
Multi-Decoder with Adversarial

- A multi-decoder model with adversarial



评价指标

- **自动评测**
  - 精确度
  - BLEU
  - G-score
- **人工评价**
  - **不同维度：**评价者需要根据情感转换度和文本语义保留程度对生成的文本进行打分



自动评测

Yelp	ACC	BLEU	G-score
CAAE	93.22	1.17	10.44
MDAL	85.65	1.64	11.85
Proposed Method	80.00	22.46	**42.38**

Amazon	ACC	BLEU	G-score
CAAE	84.19	0.56	6.87
MDAL	70.50	0.27	4.36
Proposed Method	70.37	14.06	**31.45**

人工评测

Yelp	Sentiment	Semantic	G-score
CAAE	7.67	3.87	5.45
MDAL	7.12	3.68	5.12
Proposed Method	6.99	5.08	**5.96**

Amazon	Sentiment	Semantic	G-score
CAAE	8.61	3.15	5.21
MDAL	7.93	3.22	5.05
Proposed Method	7.92	4.67	**6.08**

情感化自然语言生成

- 生成样例
- 现有方法
  - **内容经常无关**
  - 情感转换十分充分
    - 用词极为鲜明
      - Great, best, love
- Xu et al. (2018)
  - **内容保留更完整**
  - 情感转换也很充分
    - 但可能含蓄一些
      - Thanks, fresh

Input: *I would strongly advise against using this company.*
CAAE: *I love this place for a great experience here.*
MDAL: *I have been a great place was great.*
Proposed Method: *I would love using this company.*

Input: *The service was nearly non-existent and extremely rude.*
CAAE: *The best place in the best area in vegas.*
MDAL: *The food is very friendly and very good.*
Proposed Method: *The service was served and completely fresh.*

Input: *Asked for the roast beef and mushroom sub, only received roast beef.*
CAAE: *We had a great experience with.*
MDAL: *This place for a great place for a great food and best.*
Proposed Method: *Thanks for the beef and spring bbq.*

Input: *Worst cleaning job ever!*
CAAE: *Great food and great service!*
MDAL: *Great food, food!*
Proposed Method: *Excellent outstanding job ever!*

去情感实验分析

- Michael is absolutely wonderful.
- I would strongly advise against using this company.
- Horrible experience!
- Worst cleaning job ever!
- Most boring show i’ve ever been.
- Hainan chicken was really good.
- I really don’t understand all the negative reviews for this dentist.
- Smells so weird in there.
- The service was nearly non-existent and extremely rude.

错误分析

情感冲突

- Outstanding and bad service

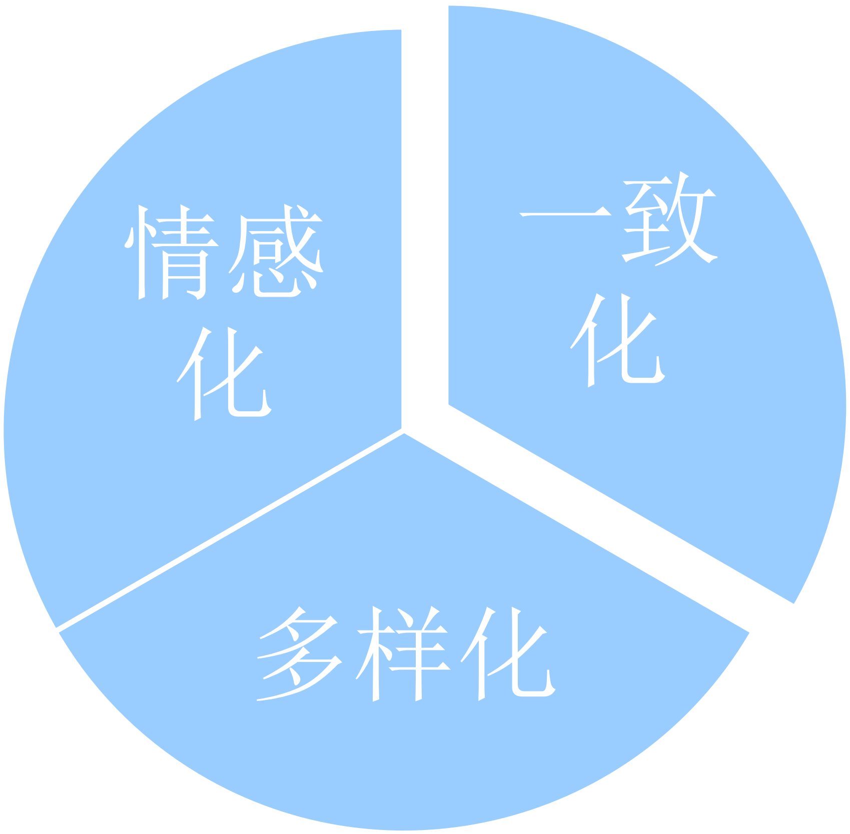
The service here is very good → Outstanding and bad service

中性句子

It’s our first time to the bar and it is totally amazing → It’s our first time to the bar

主要挑战

- 让模型说得好



如何改进—一致性

- 问题
  - 在自动文摘、文本简化等任务中编码器解码器模型会出现**生成的文本与输入文本**不一致的情况
  - 在一些任务中还会出现生成**文本内部不一致、不通顺**的情况

- 可能的原因
  - 编码器编码遗漏过多信息
  - 解码器解码不依赖编码器结果
    - Free-run的输出文本
  - 解码器解码过度依赖注意力
    - 忽视先前生成的文本
  - 等等

Text: 昨晚，中联航空成都飞北京一架航班被发现有多人吸烟。后因天气原因，飞机备降太原机场。有乘客要求重新安检，机长决定继续飞行，引起机组人员与未吸烟乘客冲突。
Last night, several people were caught to smoke on a flight of China United Airlines from Chendu to Beijing. Later the flight temporarily landed on Taiyuan Airport. Some passengers asked for a security check but were denied by the captain, which led to a collision between crew and passengers.

RNN: 中联航空机场发生爆炸致多人死亡。
China United Airlines exploded in the airport, leaving several people dead.

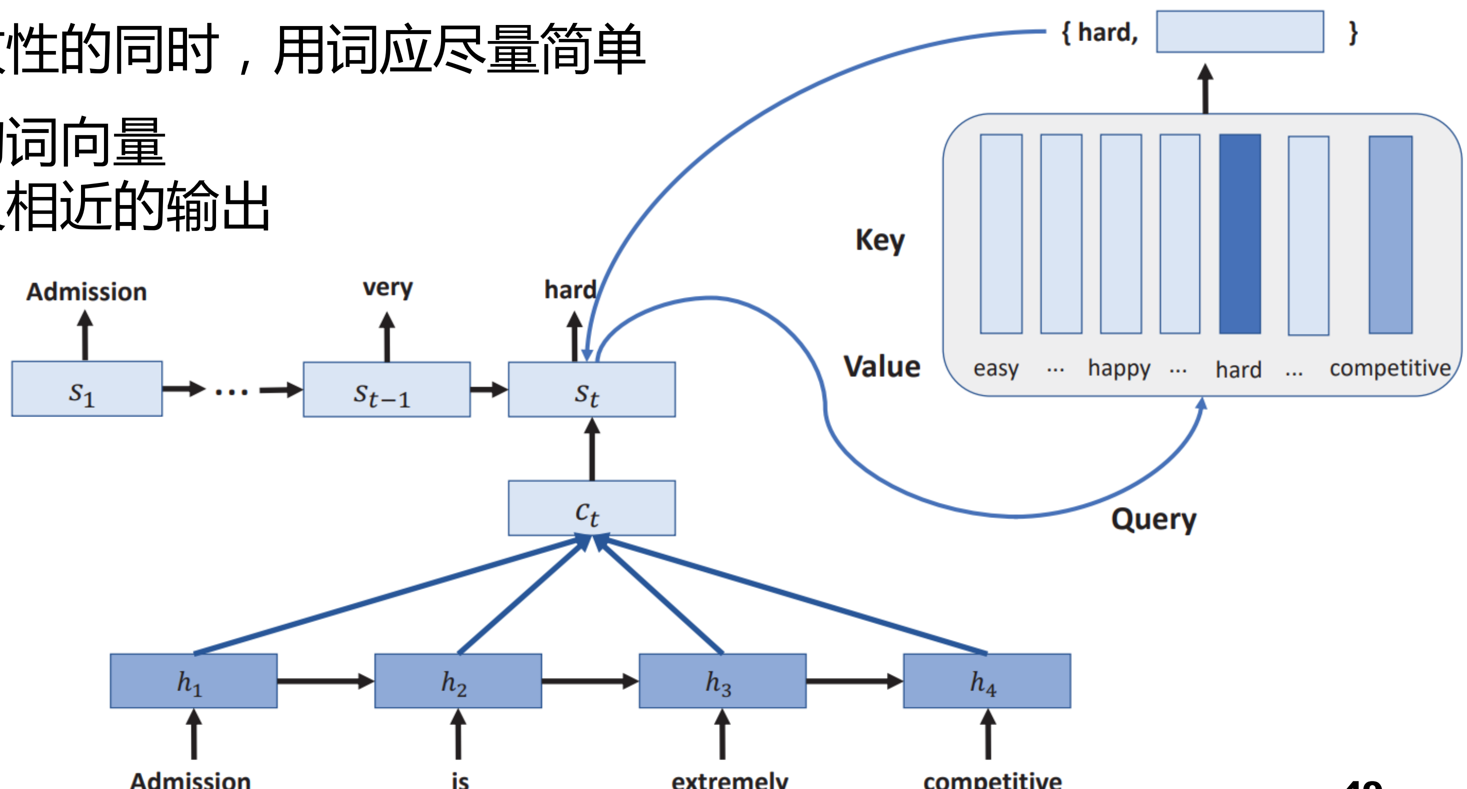
Gold: 航班多人吸烟机组人员与乘客冲突。
Several people smoked on a flight which led to a collision between crew and passengers.

样例来源：[Ma et al., 2017]

如何改进一致性

- WEAN（Word Embedding Attention Network）[Ma et al., 2018]

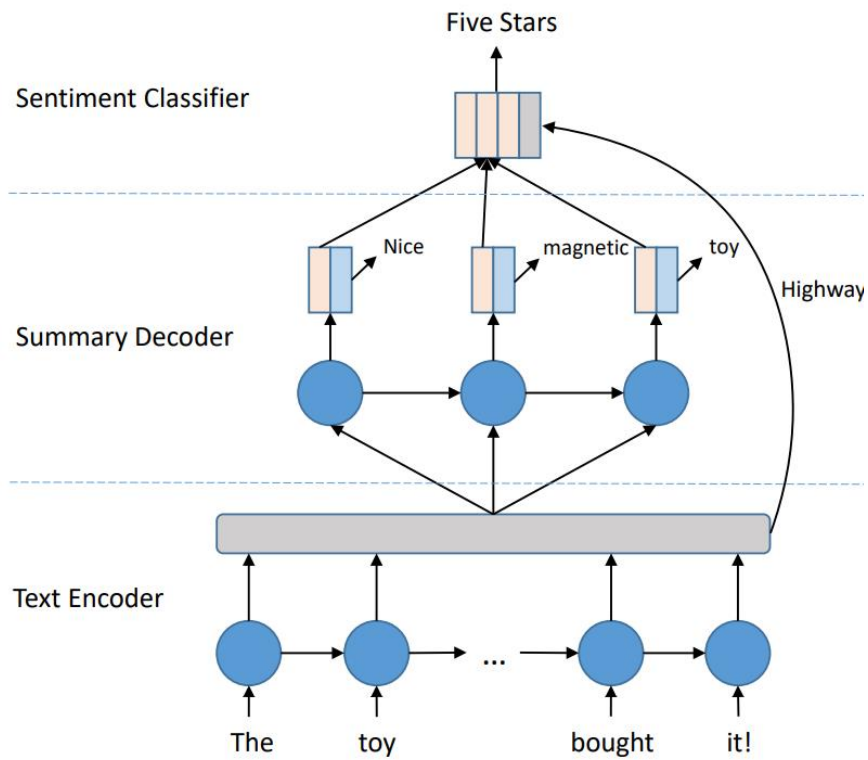
- 使文本简化真正简化文本
  - 保证一致性的同时，用词应尽量简单
  - 借助查询词向量
    获得词义相近的输出



如何改进—一致性

- **HSSC（Hierarchical Summarization and Sentiment Classification） [Ma et al., 2018]**

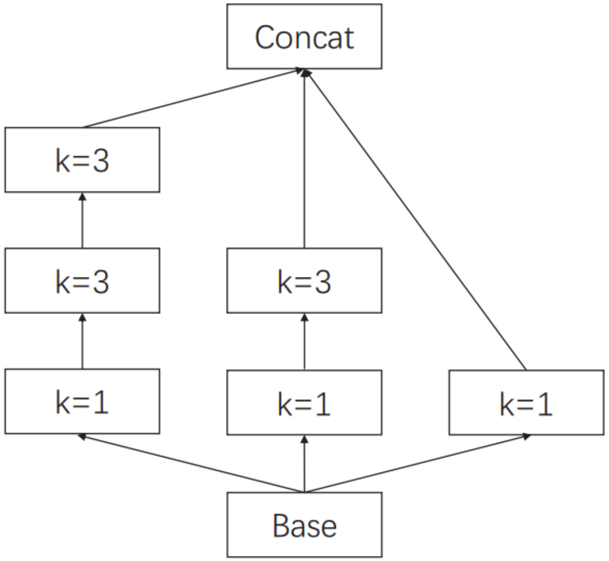
- **借助多任务学习改进与输入一致性**
  - 商品评论摘要要能反映用户整体评价
  - 引入情感分类任务作为监督
  - 相辅相成

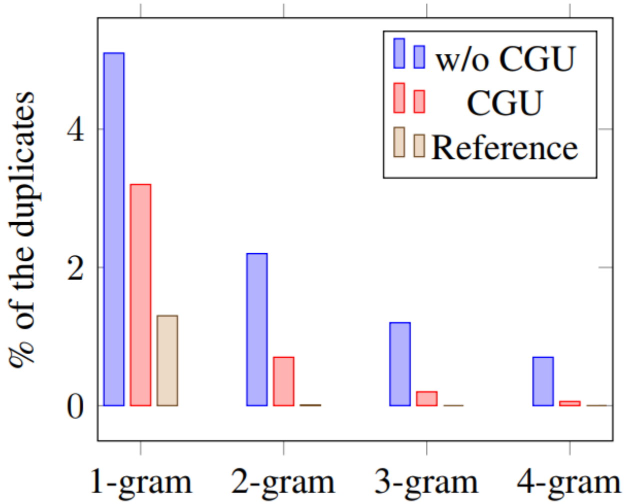


如何改进一致性

- Global Encoding [Lin et al., 2018]

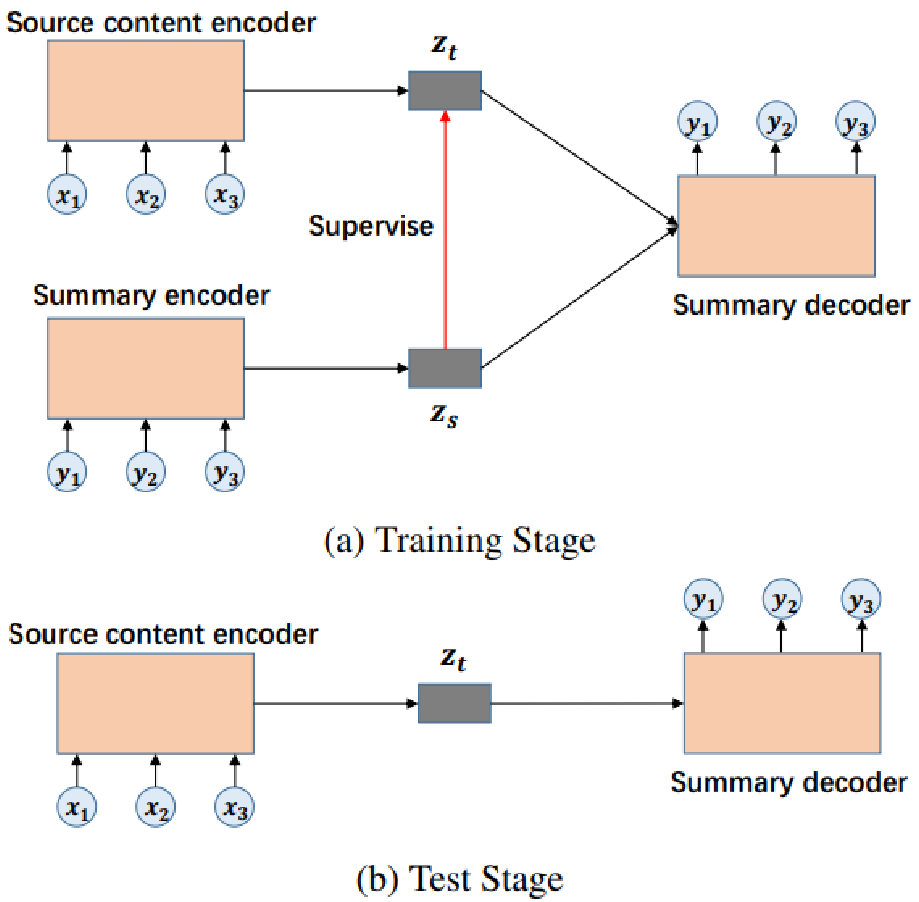
- **更好编码输入文本以提升一致性**
  - 自动文摘中的注意力会引起输出短语**重复问题**
  - 通过CNN构造输入**短语级别向量**以供注意力机制应用
  - 能够明显减少输出文本重复问题





如何改进—一致性

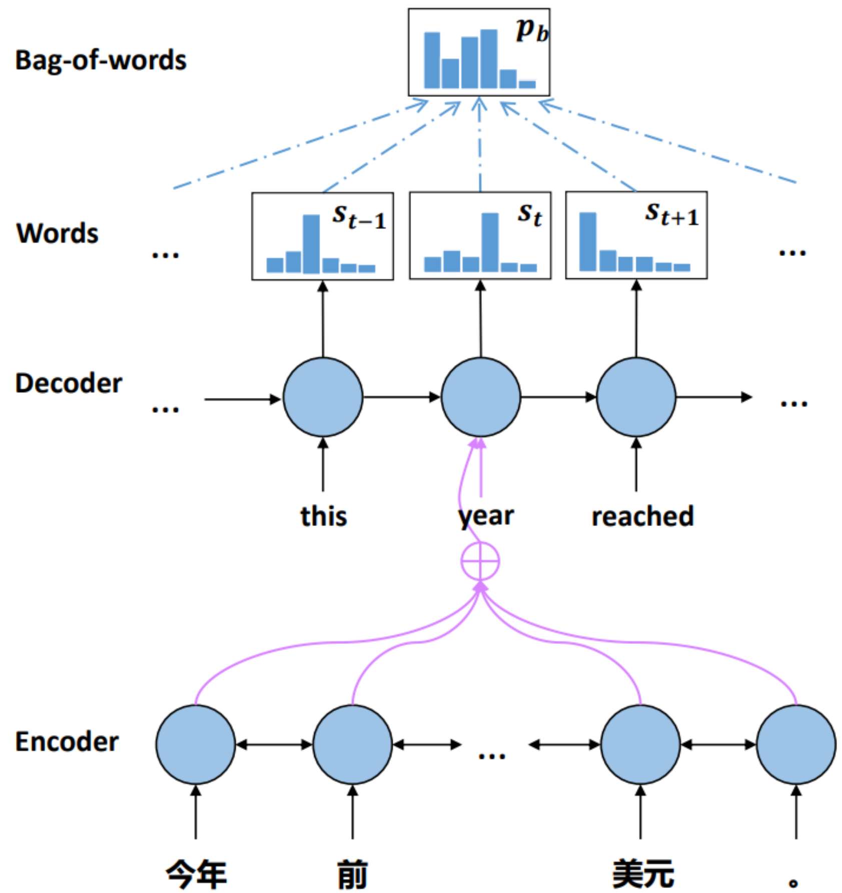
- superAE [Ma et al., 2018]
- 增强编码器编码能力
  - 摘要中输入与输出含义相近
  - 显式地在编码端剔除无关信息
    - 一个摘要的自编码器
    - 用摘要的编码向量监督原文编码
    - 传统方法编码器解码器方法没有筛选过程
  - 引入对抗学习思想
    - 编码器目标
      - 原文与摘要编码向量不可被区分



如何改进—致性

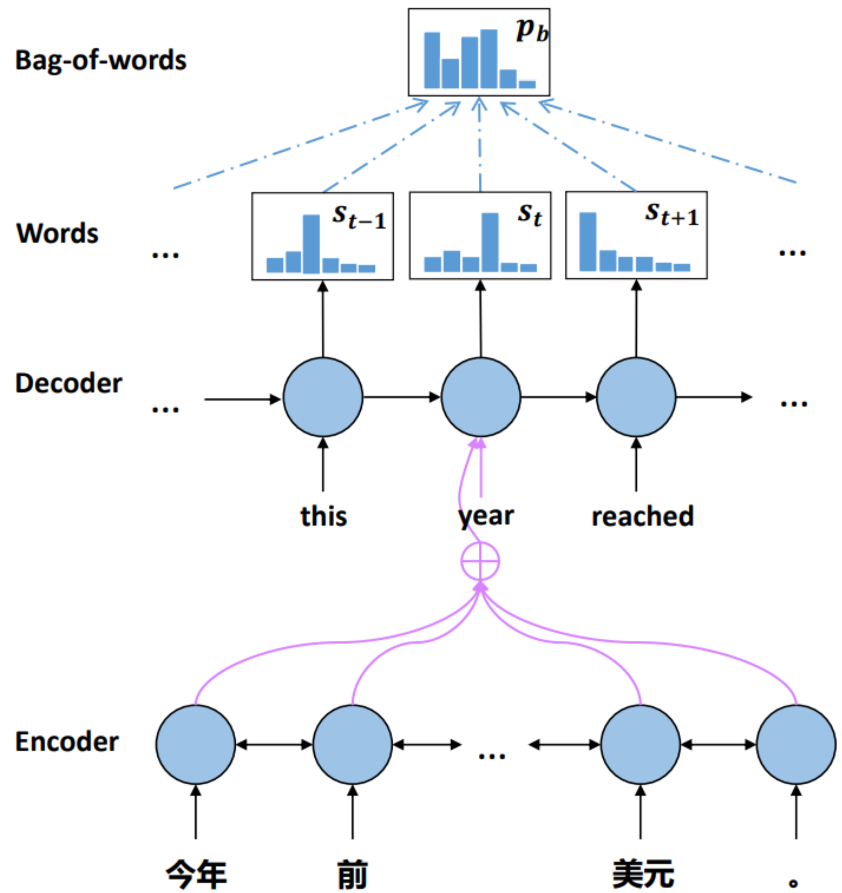
- Bag-of-Words as Target for Neural Machine Translation [Ma et al., ACL 2018]

- **改进解码器端的监督方法**
  - 原有训练要求输出端词绝对顺序相同
    - 过度惩罚那些相对次序正确的词
  - 引入新的目标函数
    - 使优化目标与实际更一致



如何改进—致性

- **Bag-of-Words as Target for Neural Machine Translation** [Ma et al., ACL 2018]

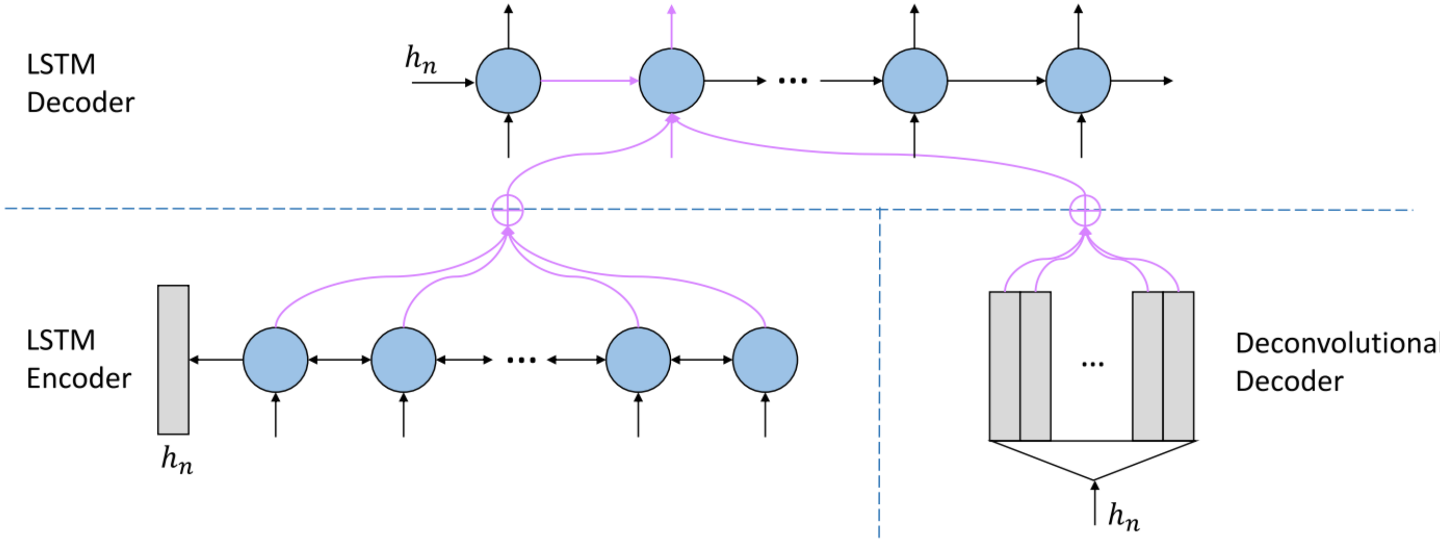


- **机器翻译：**
  句子词袋作为额外监督信号
  - 原有训练要求输出端词绝对顺序相同
    - 过度惩罚那些相对次序正确的词
  - 词袋可以容忍绝对次序错误但相对顺序正确的句子
  - 能够显著提高中英机器翻译效果

如何改进一致性

- **Deconvolution-Based Global Decoding [Lin et al., 2018]**

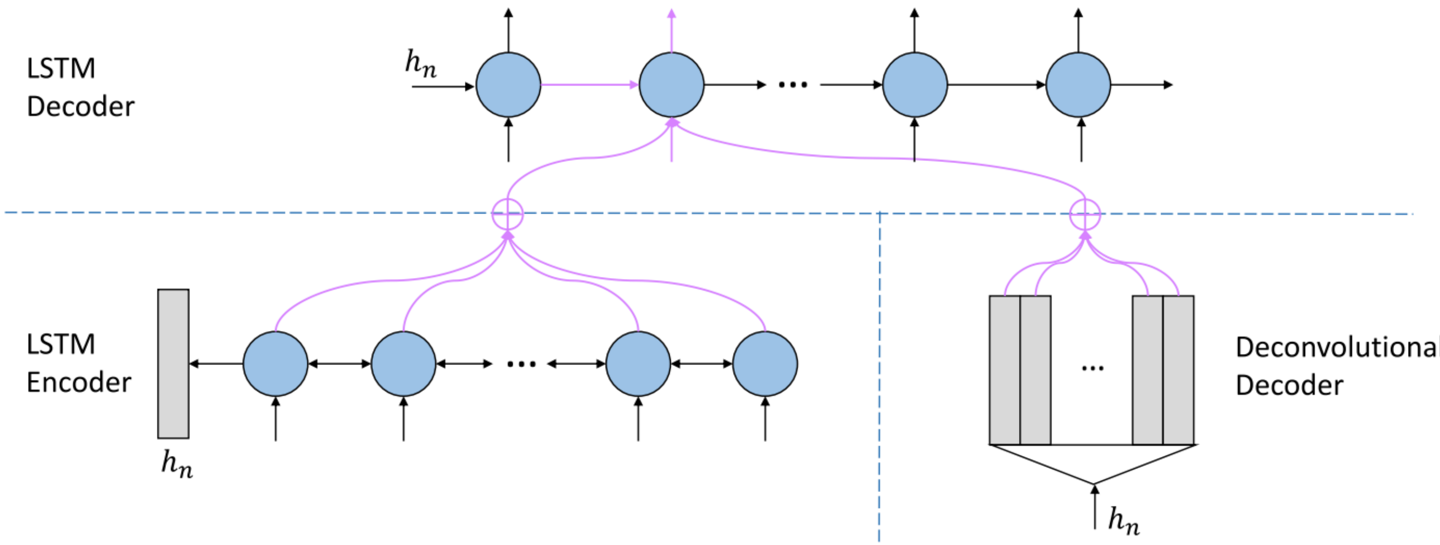
- **改进解码器以提升机器翻译效果**
  - 通过反卷积预先生成全局信息供解码器参考
  - 改进解码时内部的一致性



如何改进一致性

- Deconvolution-Based Global Decoding for Neural Machine Translation [Lin et al., COLING 2018]

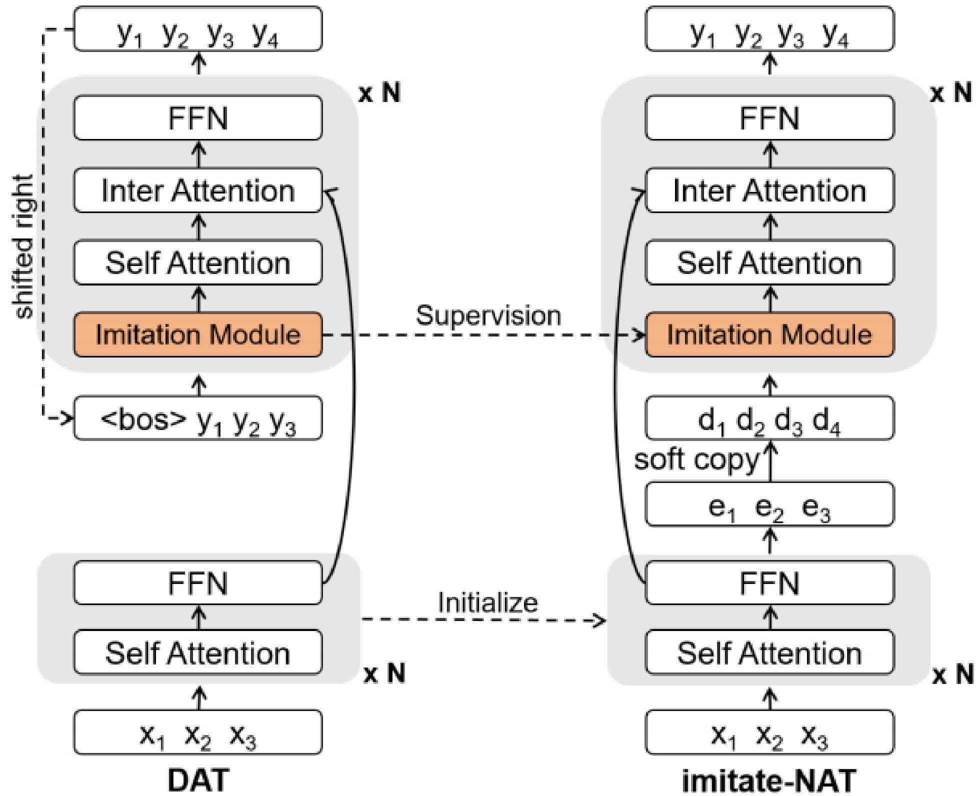
- 机器翻译：通过反卷积**预先生成全局信息**供解码器参考
  - 改进解码时内部的一致性



如何改进一致性

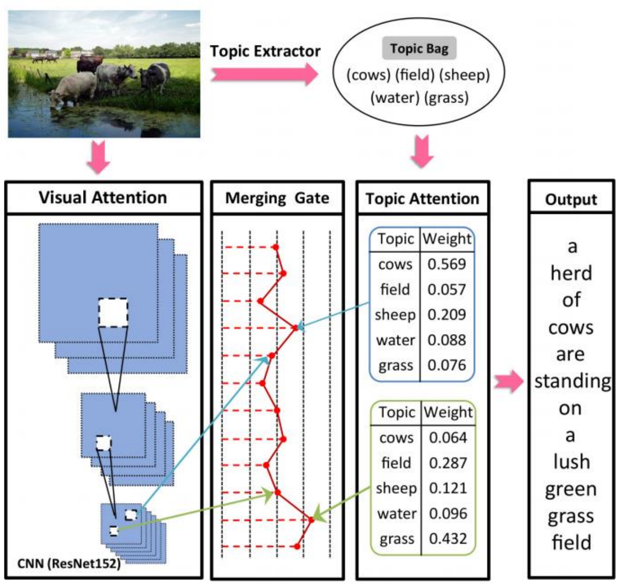
- Imitation Learning for Non-Autoregressive Neural Machine Translation [Wei et al., ACL 2019]

- **非时序机器翻译：模仿时序模型建模上下文信息**
  - 利用模仿学习，使非时序模型仿照时序模型解码，更有效的建模输出上下文，保证**内容一致性**
  - 不降低非时序模型的推理速度，同时改进翻译效果

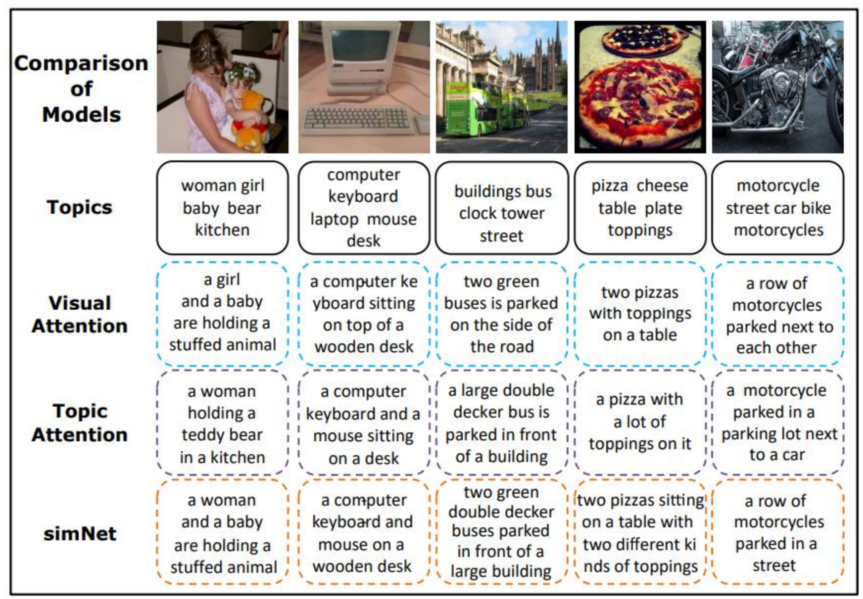


如何改进—一致性

- **simNet: Stepwise Image-Topic Merging Network for Generating Detailed and Comprehensive Image Captions** [Liu et al., EMNLP 2018]
  - 建模**图像跟文本**之间的信息一致性



图像话题融合模型

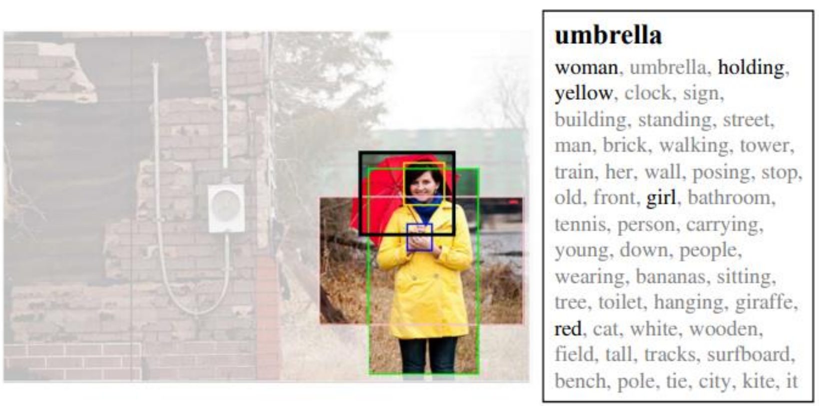


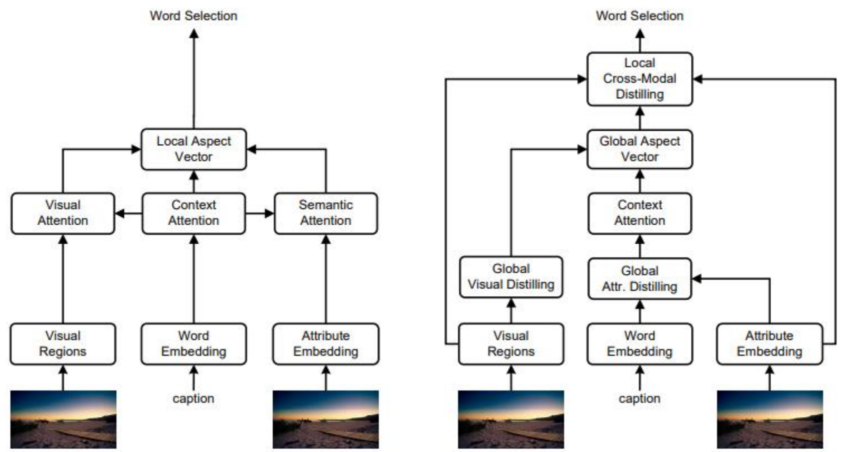
图像描述生成

Liu et al., EMNLP 2018

如何改进—一致性

- **Exploring and Distilling Cross-Modal Information for Image Captioning** [Liu et al., IJCAI 2019]
  - 利用模态内部全局信息浓缩，**提取模态内部自然关联的信息组合**
  - 利用模态间的局部信息浓缩，**提取与解码上下文匹配的源信息组合**（多框多文字组合成一个整体！）
  - 建模单一模态内部和不同模态之间的**信息一致性**





捕捉模态内部和模态间自然关联的信息

基于Transformer的跨模态信息浓缩模型

THANK YOU

Questions?