

基于预训练语言模型的可控文本生成的研究与应用

澜舟科技是一家认知智能公司，针对商业场景数字化转型、以自然语言处理为基础提供商业洞见类产品。

主讲人：周明，澜舟科技创始人&CEO

中国计算机学会副理事长

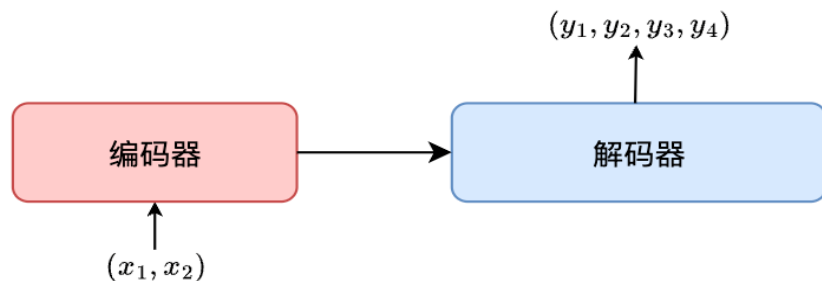
原ACL主席、微软亚洲研究院副院长

zhouming@langboat.com

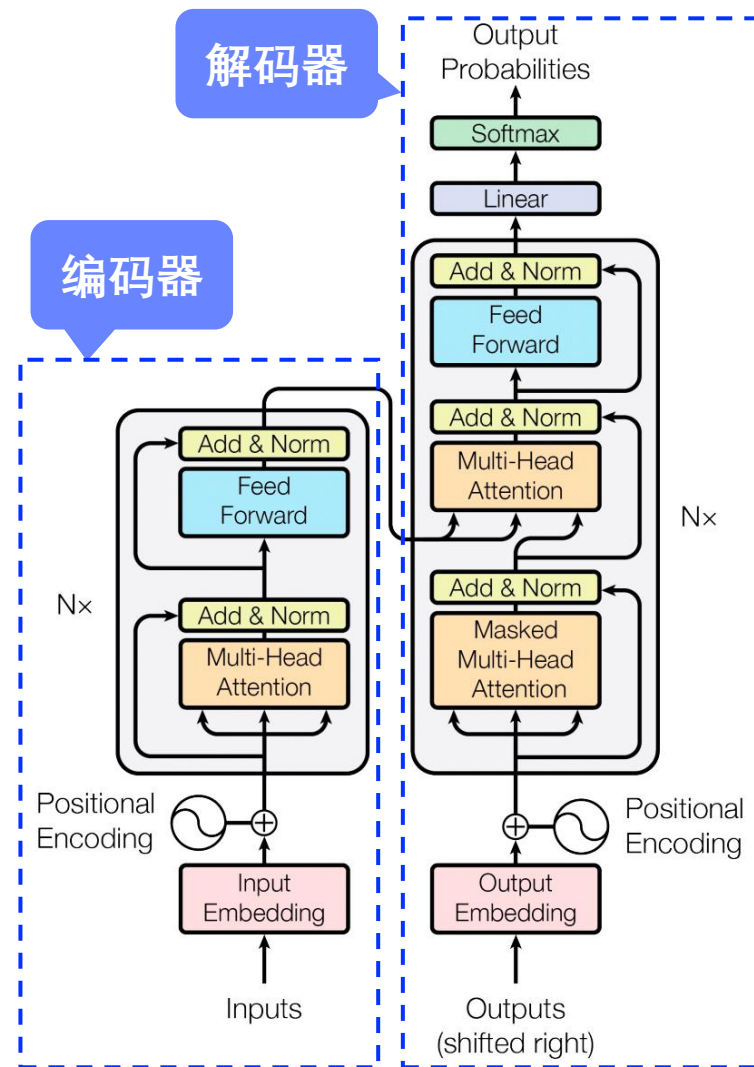
- 01** 文本生成基本方法与应用
- 02** 文本生成中的可控方法研究
- 03** 文本生成中融入知识、常识的方法研究
- 04** 长文本生成方法研究
- 05** 文本生成中的解码方法研究

文本生成任务及主流框架

- 文本生成的目标是生成自然语言文本，包括多种形式的输入
 - 输入结构化数据的 Data-to-text
 - 输入图片数据的 Image caption
 - 输入文本数据的 Text-to-Text 任务，包括对话生成、故事生成、文案生成等
- 主流文本生成模型
 - 基于 Transformer 的编码器-解码器框架



编码器-解码器框架



基于Transformer的编码器-解码器框架

- 可控文本生成旨在使生成的文本带有所期望的某种属性，如情感、关键词、主题、事实等

情感

前文：这首歌_
情感：积极

这首歌充满感情、充满正能量、非常好听！

关键词

自行车 数显屏 骑行 健身 体型
舒适

自行车主要成分是精钢,骑行时,自行车的体型和重量都会发生变化,所以骑行自行车时一定要选择一款舒适、轻便的自行车。数显屏,可以实时显示骑行数据,健身时也能一目了然。

知识图谱

<帆布鞋, 材质, 织物> <帆布鞋, 功能, 透气、耐磨> <帆布鞋, 风格, 休闲>

帆布鞋的材质是织物的,所以透气耐磨的功能也是比较强的。在休闲风格中,帆布鞋的功能也是比较多的。

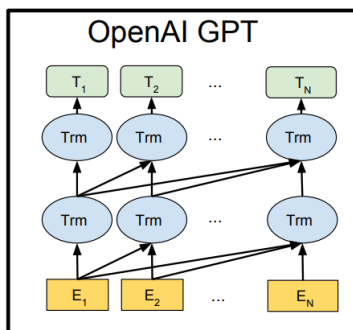
键值对

自我描述:学设计学生|情感:不用担心|品类:蓝牙耳机,笔记本

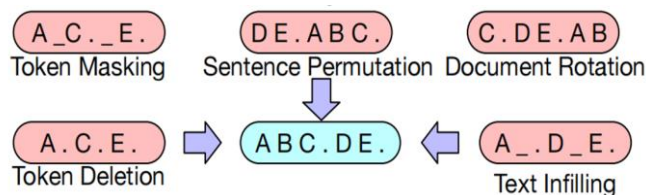
戴尔蓝牙耳机,你值得拥有!作为学设计的学生,我们对于产品的需求就是满足学生的各种需求。最近我入手了一款戴尔笔记本的蓝牙耳机~它的外形和外观都是简约大气的设计感~再加上其出色的外观设计与做工工艺让我丝毫不用担心哦,而且这个音质表现还是非常不错的,这款是支持杜比全景声效设计的耳机。

预训练语言模型极大推进了文本生成的研究与落地——GPT(自回归-解码器)、BART(自编码-编码解码器)、T5(多任务-编码解码器)等

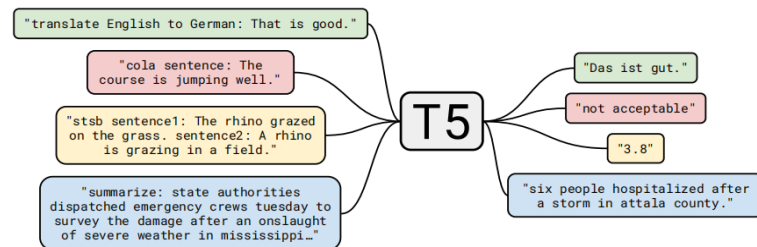
- 基于Transformer神经网络模型，语言特征提取能力强
- 自监督学习目标，能够从大量无标注数据学习丰富的语言知识
- 多任务、跨语言联合建模，模型通用性强



GPT: 语言模型学习目标



BART: 降噪自监督学习目标



T5: 多任务联合建模

问题类型	示例
常识错误	今天天气晴朗，小明没有伞被雨淋湿了，受凉感冒了。“这样啊……那就好办了!我们去医院看看吧!小明感冒了需要治疗的话一定不能耽误太久的时间，否则会影响到孩子身体健的。”“嗯嗯，你说得对，我也是这么想的，所以我们现在就出发吧!
内容逻辑错误	香水瓶也有高颜值，香味持久，干净清爽，留香时间长，不会有残留的味道，橘子的清香，闻起来很舒服…… 最后就是洗手液和护手霜啦，这两款都很平价，但是功效都不一样，有时我会用它们来卸除防晒，效果也非常的好……
内容发散	分享一下我最近爱用的身体素颜霜和香水!! 都是平价又好用，学生党也可以闭眼入的东西! .1.hfp钻石霜这个是我在李佳琦直播间种草的，我买的是01号色，刚开始我以为是粉底液，用了之后发现是真香! 不会假白，肤色提亮效果很明显，而且不蹭衣服，很自然! 我一般晚间护肤的时候用它……
语句重复	太绝了我又回购了莲木叙零食这次是核桃枣泥糕和美式牛肉绝了! 核桃枣泥块可以吃好多天了还是核桃肉我又回购了核桃肉枣泥块真太绝了太绝了! 核桃肉枣泥块可以吃好多天了还是核桃肉枣泥块太绝了! ……

- 如何更好实现生成文本的可控性
 - 可控文本生成模型构建方法
- 如何提高生成文本的事实正确性
 - 文本生成中融入常识、知识的方法
- 如何改进生成文本的前后一致性、连贯性
 - 长文本生成建模方法
- 如何改进重复生成，增加多样性
 - 文本生成模型的训练与解码方法

01 文本生成基本方法与应用

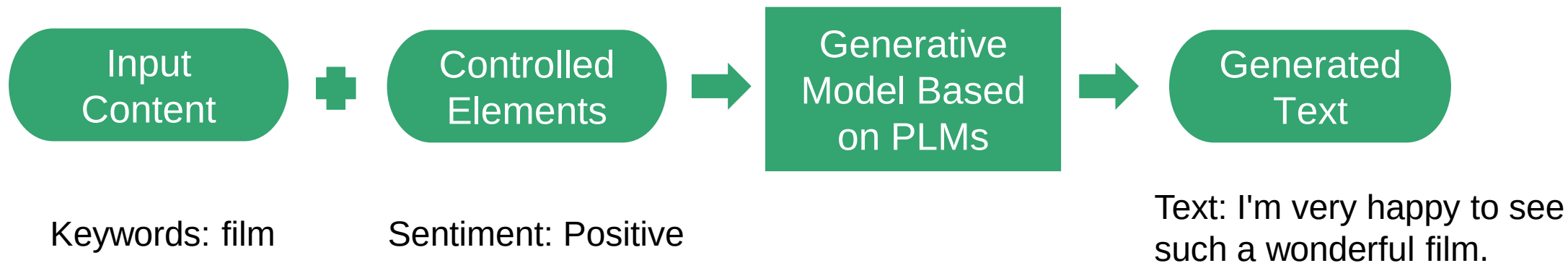
02 文本生成中的可控方法研究

03 文本生成中融入常识、知识的方法研究

04 长文本生成方法研究

05 文本生成中的解码方法研究

- 调整解码策略 (Decoding strategies based method)
 - 使生成结果尽可能包含目标内容 (如主题关键词)
- 调整训练目标函数 (Training objectives based method)
 - 构建面向特定任务的可控训练目标函数
- 调整模型输入 (Prompting-based method)
 - 通过输入控制元素影响生成结果



Muhammad Khalifa, Hady Elsahar, and Marc Dymetman. 2021. A distributional approach to controlled text generation. ICLR.

Damian Pascual, Beni Egressy, Clara Meister, Ryan Cotterell, and Roger Wattenhofer. 2021. A plugand-play method for controlled text generation. EMNLP.

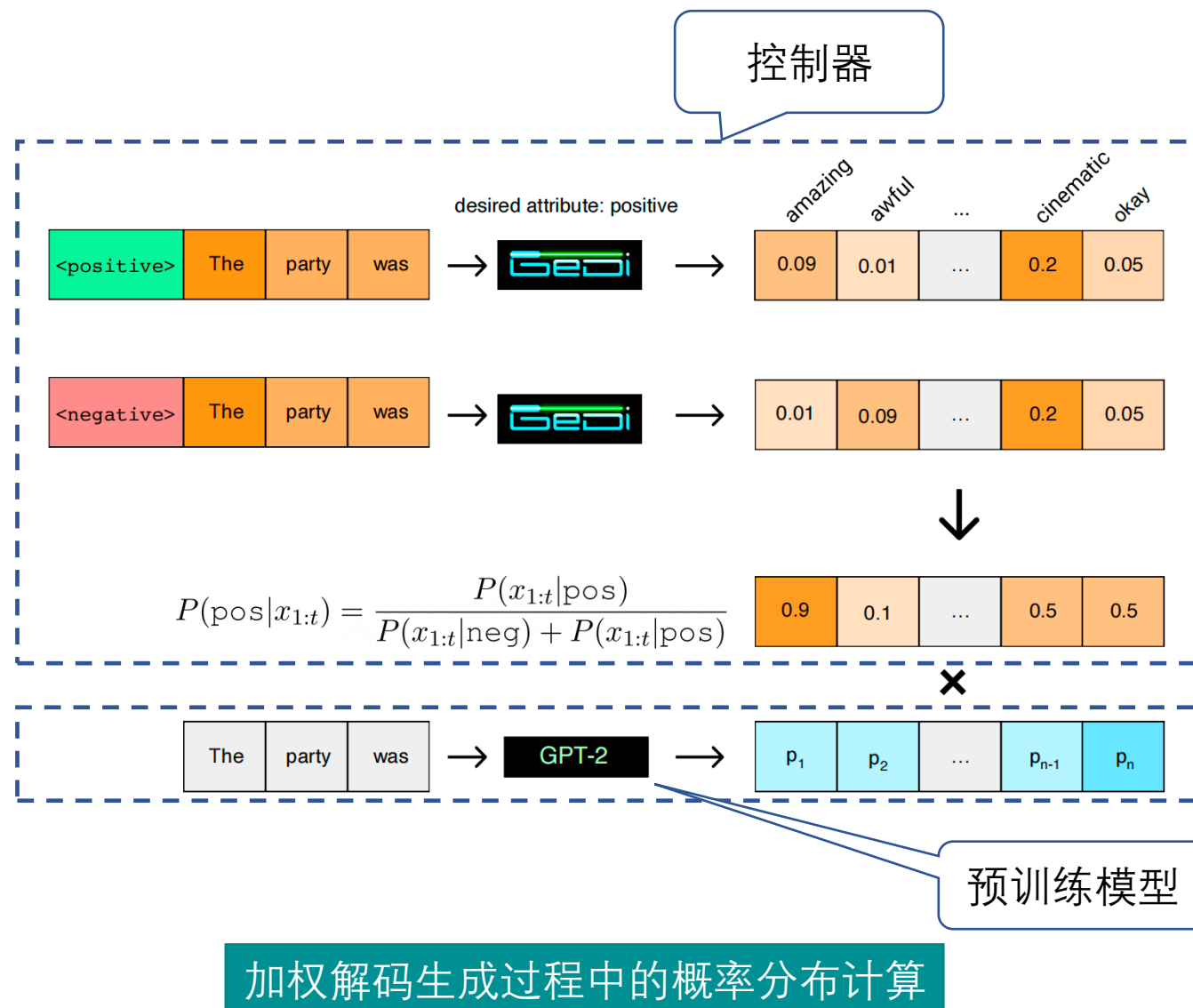
Jordan Clive, Kris Cao, and Marek Rei. 2021. Control prefixes for text generation. arXiv preprint arXiv:2110.08329.

- 可控训练方法存在的问题
 - 随着预训练模型的参数量越来越大，对预训练模型进行可控训练需要的计算成本也越来越高
- 改进方法
 - 固定预训练语言模型参数，解码阶段引入轻量化的控制器来调整语言模型的输出概率，鼓励模型生成符合控制要素的内容，抑制不符合的内容

条件概率计算方法

$$P(X|a) \propto \prod_{i=1}^n \left[P(x_i|x_{<i}) P(a|x_{<i})^\lambda \right]$$

控制器



- 上述工作存在的问题
 - 计算条件概率过程中赋予控制器的权重 λ 是固定不变的，不能根据语言模型当前的解码状态动态调整，当控制器影响较大时，容易产生不流畅的文本

$$P(X|a) \propto \prod_{i=1}^n [P(x_i|x_{<i})P(a|x_{<i})^\lambda]$$

λ 表示控制器的权重

- 改进方法
 - 增加调节器(Regulator)，根据语言模型解码序列动态调整（缩小或放大）控制器的影响

调节器计算方法

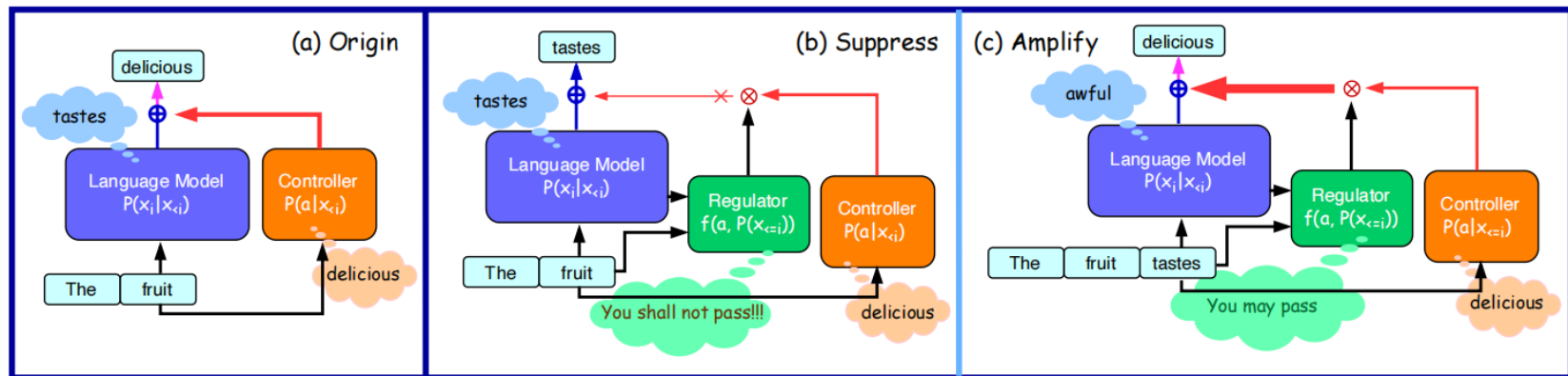
$$W^a = \{w_1, w_2, \dots, w_k\}$$

$$t_H = \sum_{w \in W^a} P(x_i = w | x_{<i})$$

$$f = f_H(W^a, P(x_i | x_{<i})) \\ = t_H / \tau_H,$$

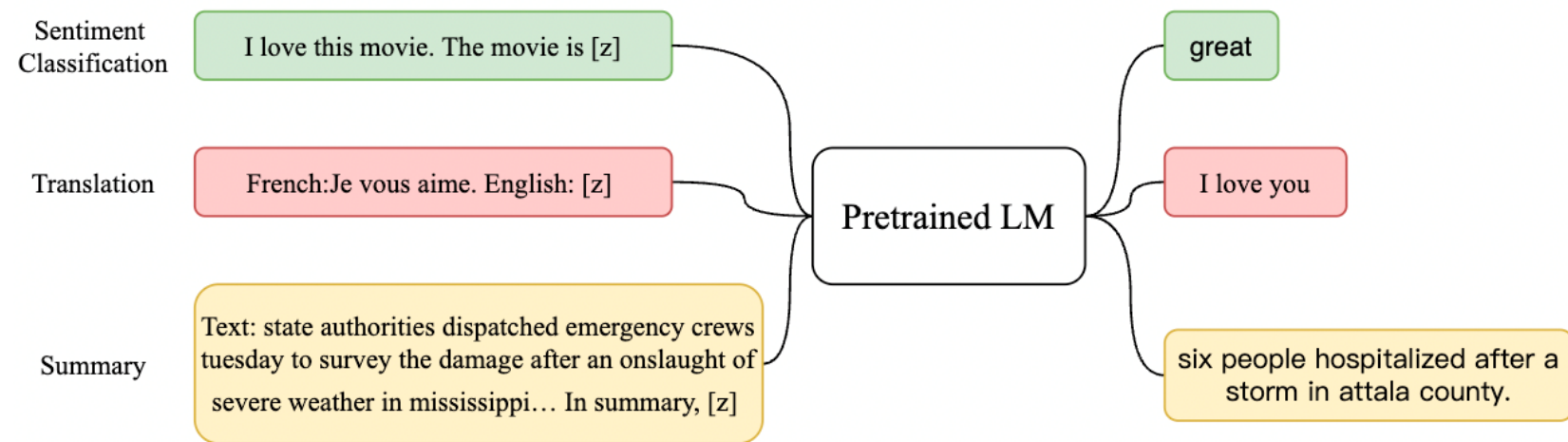
$$P(X|a) \propto \prod_{i=1}^n [P(x_i|x_{<i})P(a|x_{<i})^{\lambda f(a, P(x_{\leq i}))}]$$

调节器动态更新控制器的权重



基于prompt的可控文本生成方法

- 实现方法及优势
 - 通过设计适当的“prompt”，控制模型的输出，使预训练语言模型可以基于“prompt”预测出所需的答案
 - 优势是给定一组合适的 prompt，以及单个预训练语言模型就能够解决多个任务，从而实现目标，如可控文本生成



不同任务下的prompt, [z]表示需要输出的文本

这首歌充满感情、充满正能量、非常好听！它表达了……

Pretrained LM

这首歌充满感情

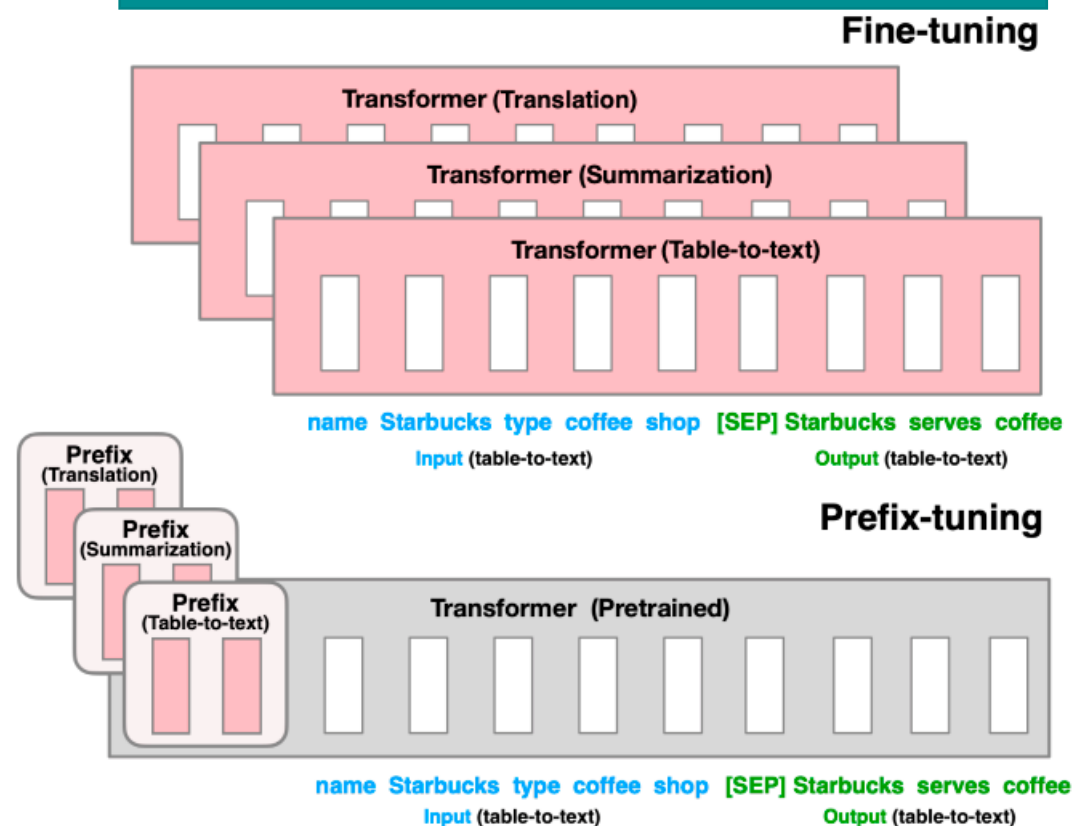
基于prompt的文本续写

基于连续prompt向量的可控文本生成方法

- 早期的prompt方法
 - 主要通过人工制作prompt，需要大量的时间和经验
 - Prompt和其他输入共享预训练语言模型参数，不易进行下游任务的适配

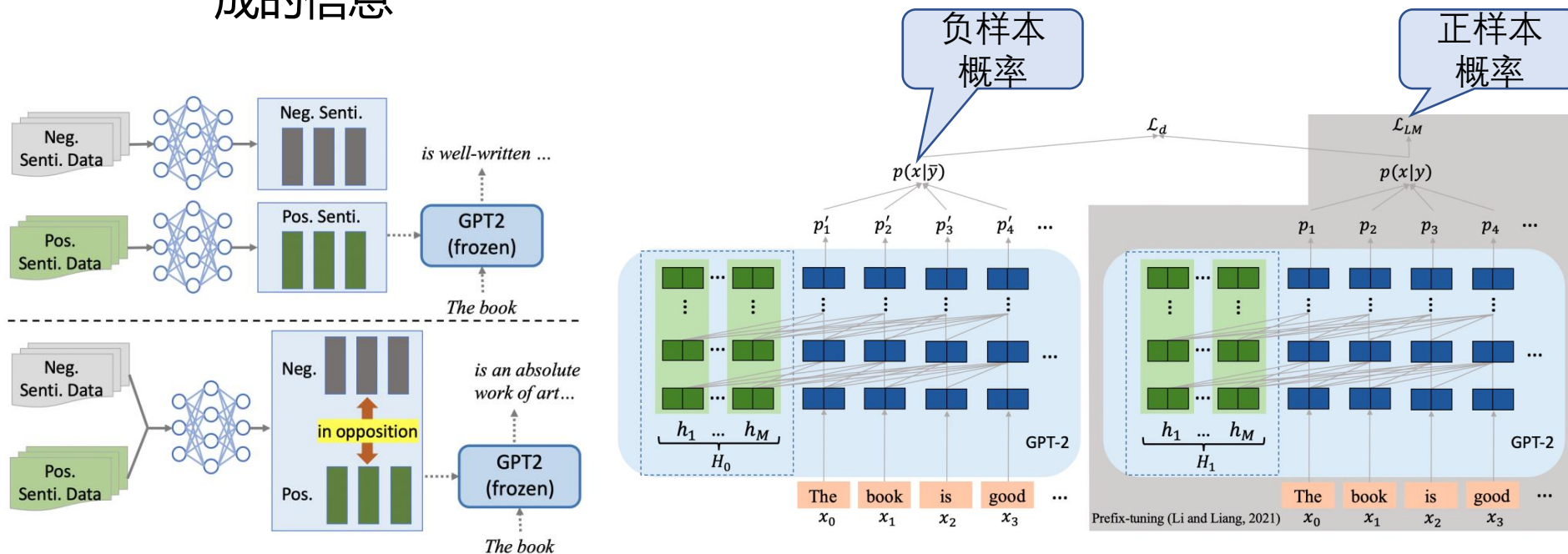
- 改进方法
 - 将具体的prompt替换为一小段任务特定的连续向量（prefix）
 - 微调阶段固定预训练模型的参数，只优化prefix向量
 - 引入prefix向量并在特定任务上微调，避免了人工制作prompt

全量参数微调和prefix参数微调方法的对比



基于对比前缀的可控文本生成方法

- 对于属性可控的生成任务，如情感可控（积极、消极）
 - 先前工作将积极或消极属性视为独立的控制任务，分别训练每个任务的prefix向量，忽视了属性之间的关系
- 改进方法：
 - 考虑属性之间的对立关系，如积极和消极的对立
 - 采用对比损失，同时训练多个prefix向量，向prefix向量注入鼓励生成的信息和不鼓励生成的信息



$$\mathcal{L}_{sup} = \omega_1 \mathcal{L}_{LM} + \omega_2 \mathcal{L}_d \quad (1)$$

$$\mathcal{L}_{LM} = - \sum_{t=1}^T \log p(x_t | x_{<t}, y) \quad (2)$$

$$\mathcal{L}_d = - \log \frac{p(y)p(x|y)}{\sum_{y' \in Y} p(y')p(x|y')} \quad (3)$$

各个属性独立学习的方法

多属性联合学习的方法

- 可持续学习（Continual learning）的意义和挑战
 - 现实世界中需要学习新任务总是不断出现的
 - 当学习新任务时，模型容易忘记原先已经学习到的知识
 - 面对多个任务时，需要多个不同模型，训练和部署代价高昂
 - 需要有效的学习机制，实现跨任务知识的有效迁移
- 改进方法
 - 引入自适应的组合模块（Adaptive Compositional Modules），保证模型的持续学习能力和知识迁移能力
 - 采用轻量化微调技术，提高训练效率

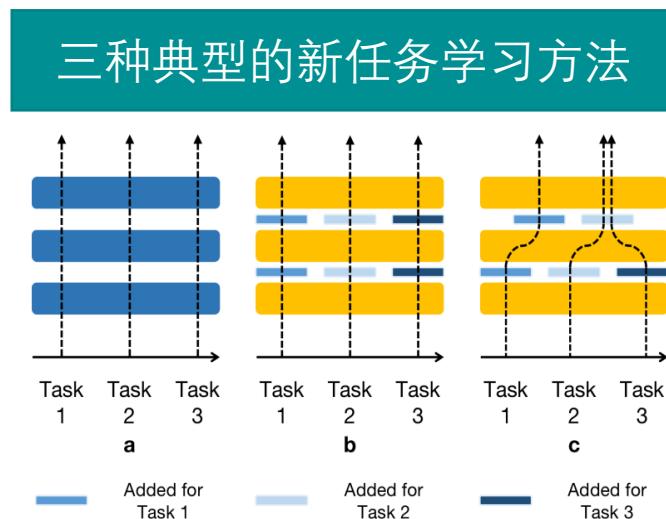


Figure 1: Comparison between previous methods (a and b) and our proposed method (c), from a multi-layer transformer model perspective. The blue blocks refer to learnable modules and the yellow blocks refer to frozen pretrained modules. **a**: retrain the whole model every time when new tasks arrive. **b**: insert task-specific modules for each task, while keeping the pre-trained model frozen. **c**: detect reusable old modules and add new modules adaptively.

- 模型结构决定阶段
 - 决定复用模块，决定新增模块
- 新任务模型训练阶段
 - 确定模型结构后，利用新数据训练模型

基于轻量化参数模块的持续学习

各个旧任务参数的概率选择

$$\lambda_{i,l} = \frac{e^{c_{i,l}}}{\sum_{t=1}^{k+1} e^{c_{t,l}}}, i = 1 \dots k + 1$$

根据概率选择决定参数利用方法

$$h_{mh}^l = \sum_{t=1}^{k+1} \lambda_{t,l} MLP_{MH}^{t,l}(o_{mh}^l)$$

$$h_{ff}^l = \sum_{t=1}^{k+1} \lambda_{t,l} MLP_{FF}^{t,l}(o_{ff}^l)$$

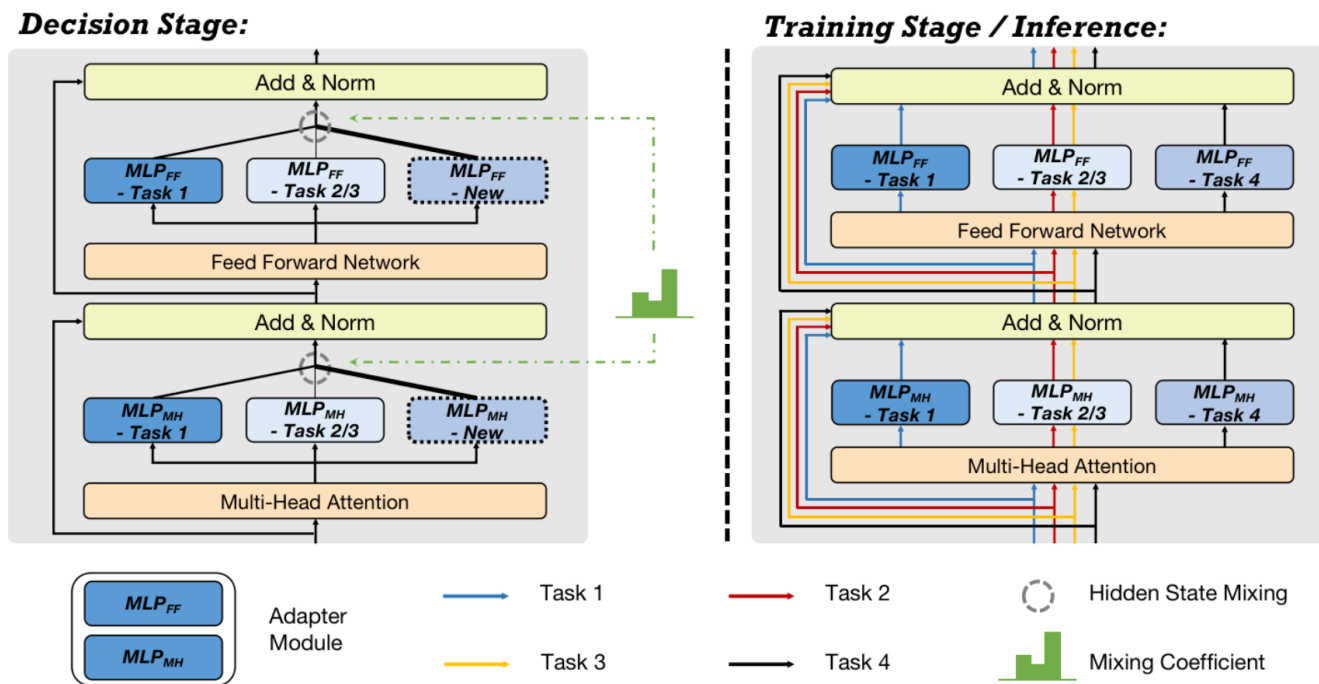


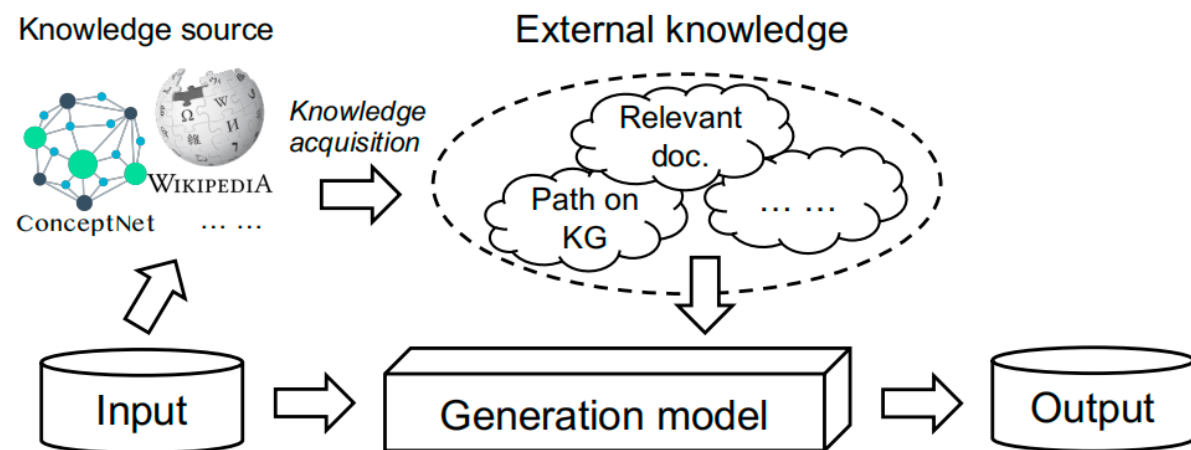
Figure 2: Our proposed model architecture with adaptive compositional modules for transformer layers. Assume after learning three tasks (1, 2, 3), we have one module for task 1, and another for task 2 and 3 in this layer. **Left:** During decision stage for task 4, we first insert a new module at this position, then all inserted modules will be used for selection using hidden state mixing. **Right:** Assume that we finally decide to add one module at this position, then each task would use its own architecture during training stage and inference.

- 01 文本生成基本方法与应用
- 02 文本生成中的可控方法研究
- 03 文本生成中融入常识、知识的方法研究**
- 04 长文本生成方法研究
- 05 文本生成中的解码方法研究

- 人类世界中广泛存在各种类型的知识，如常识知识、事实知识等
- 现有神经文本生成模型缺乏对知识的理解和建模，其生成内容存在逻辑性问题，违背事实/常识等问题
- 引入知识能帮助神经文本生成模型理解输入信息，从而生成高质量内容

Common sense Knowledge	(telephone, UsedFor, communication)
	(cooks spaghetti, Intent, to eat)
Factual Knowledge	(American, President, Joseph Robinette Biden Jr.)
	(China, GDP rank, second in the world)

常识、知识的一般表示方法



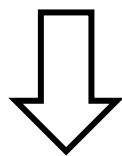
文本生成中融合常识、知识的流程

隐式地融合常识、知识的方法

- 将结构化的常识知识转化为自然语言文本的形式
- 利用这些自然语言文本训练语言模型，将常识知识注入到模型参数中，增强语言模型对知识的理解

常识-结构化三元组形式

Knowledge Bases	Original Triples
ConceptNet	(eiffel tower, AtLocation , paris) (telephone, UsedFor , communication)
ATOMIC	(PersonX dates for years, oEffect , continue dating) (PersonX cooks spaghetti, xIntent , to eat)

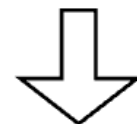


常识-自然语言形式

Knowledge Bases	Examples of Transformed Sentences
ConceptNet	eiffel tower is at paris. telephone is used for communication.
ATOMIC	PersonX dates for years. PersonY will continue dating. PersonX cooks spaghetti. PersonX wants to eat.

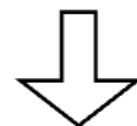
常识、知识融入语言模型

(eiffel tower, **AtLocation**, paris)
(telephone, **UsedFor**, communication)



triple → sentence

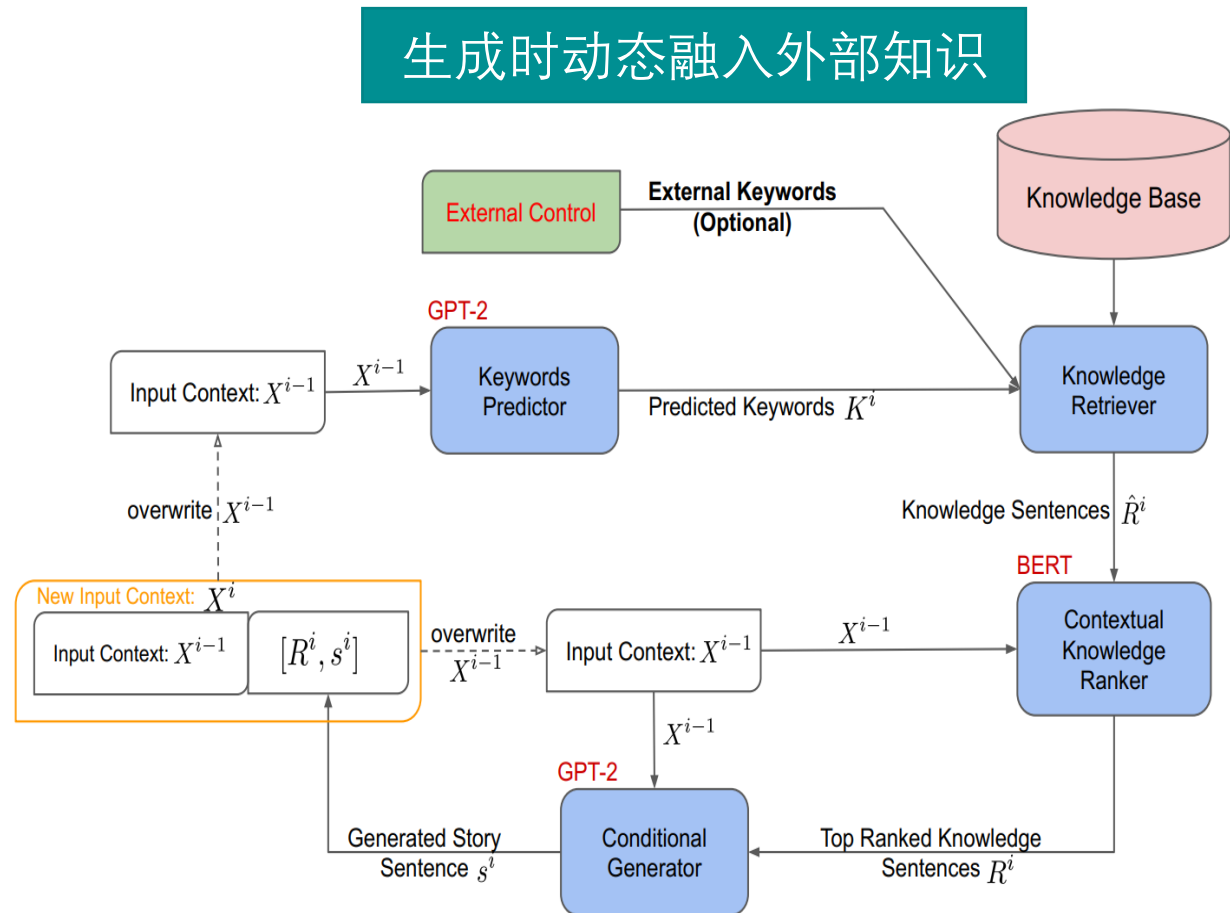
eiffel tower **is at** paris.
telephone **is used for** communication.



Continue train

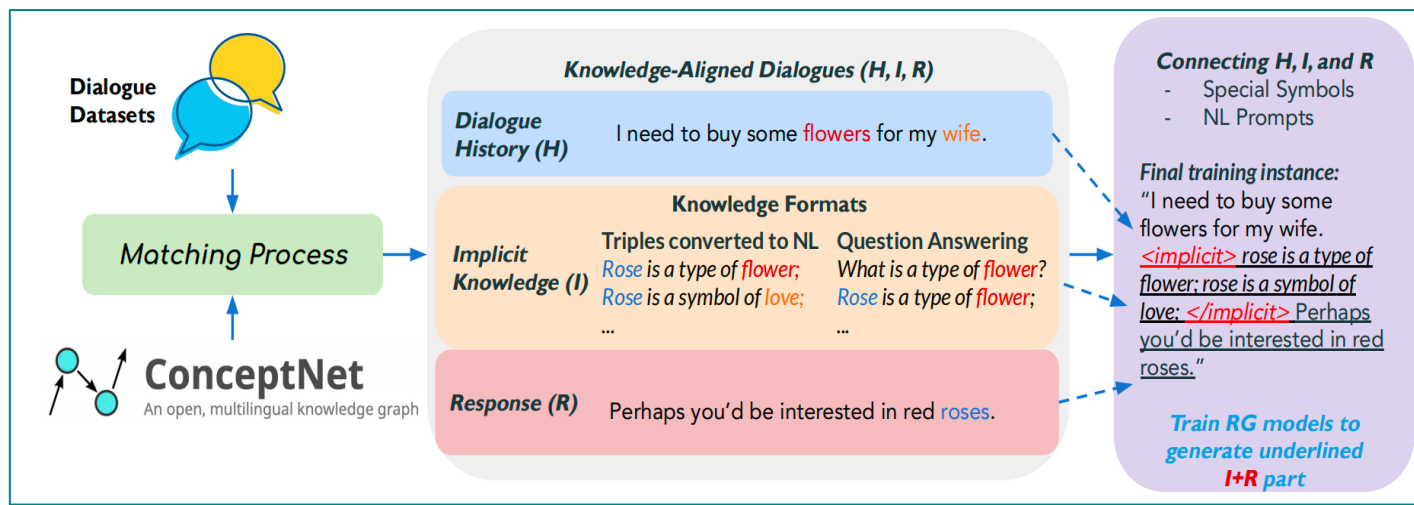
GPT2

- 隐式融合方法存在的缺点是不能准确获取生成文本使用了哪些知识，可解释性差
- 基于检索的方法，根据输入上文，从常识知识库中检索相关的常识三元组，引导模型生成符合逻辑的文本
- 生成阶段包括3个步骤：
 - 关键词预测：根据上文，预测未来关键词
 - 常识、知识检索：根据预测的关键词，从常识、知识库中检索相关常识、知识
 - 文本生成：根据上文和常识生成文本

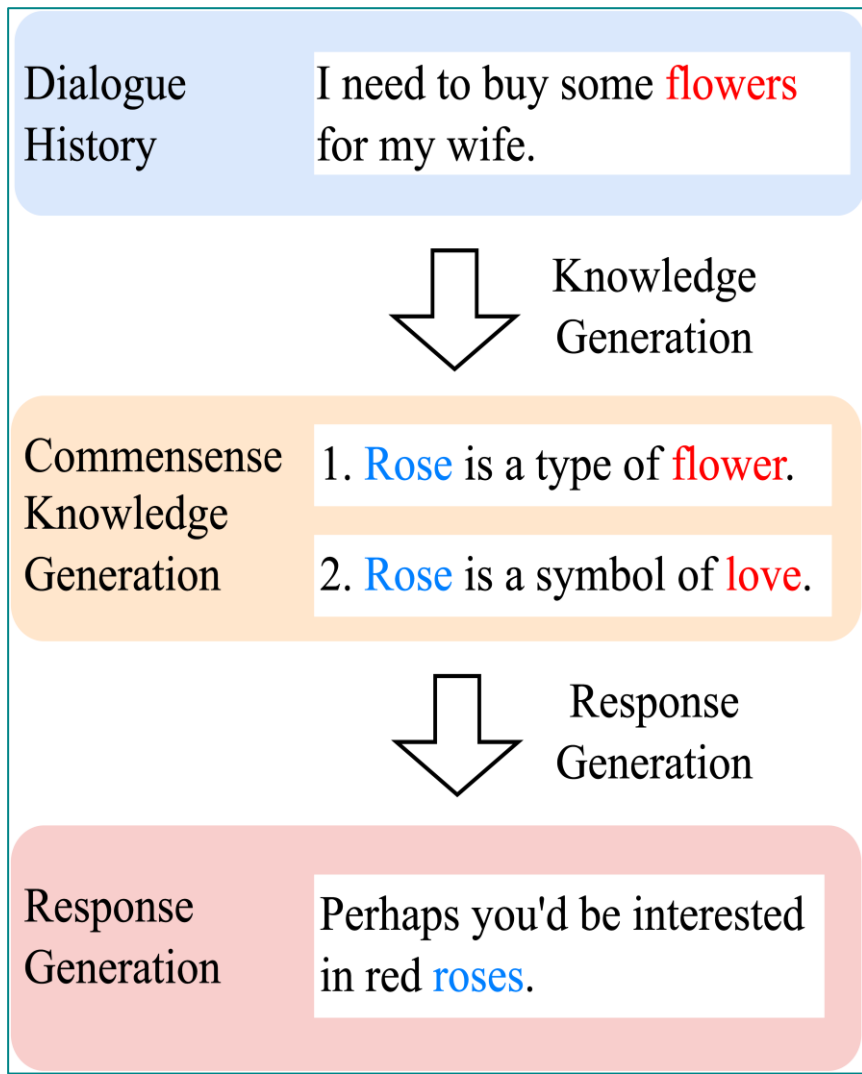


动态生成常识并融入生成的方法

- 基于检索的知识获取方法存在的问题：
 - 检索阶段计算成本高，过程复杂，且生成质量受知识库覆盖范围的影响大，模型泛化能力弱
- 改进方法：
 - 提出基于生成式的常识获取方法
 - 输入历史信息，**知识生成模型**生成相关常识知识，**文本生成模型**根据常识知识生成符合逻辑的文本



构建知识对齐的对话数据，训练知识/文本生成模型



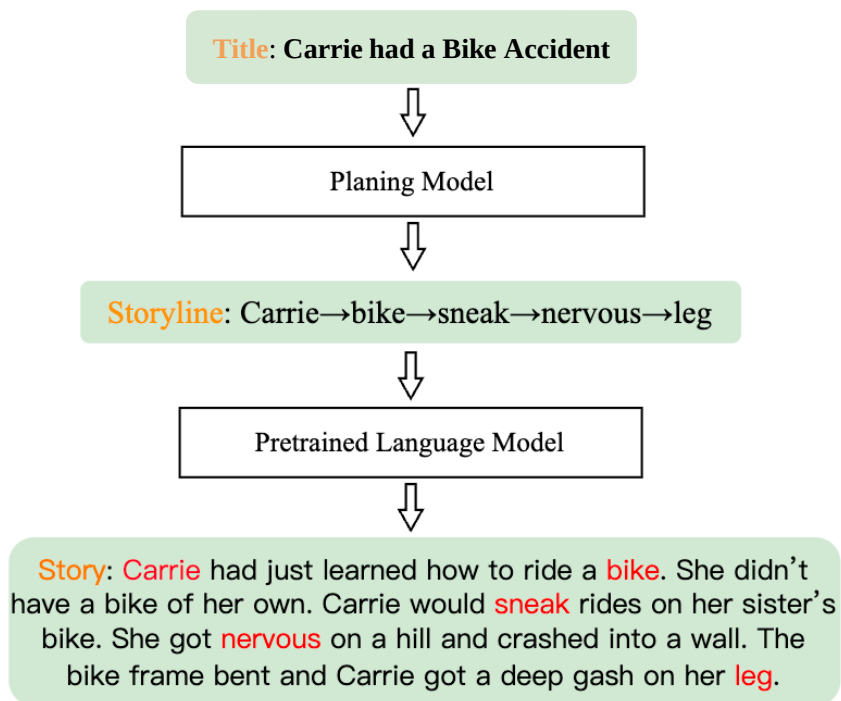
生成常识知识指导回复生成流程

- 01 文本生成基本方法与应用
- 02 文本生成中的可控方法研究
- 03 文本生成中融入常识、知识的方法研究
- 04 长文本生成方法研究**
- 05 文本生成中的解码方法研究

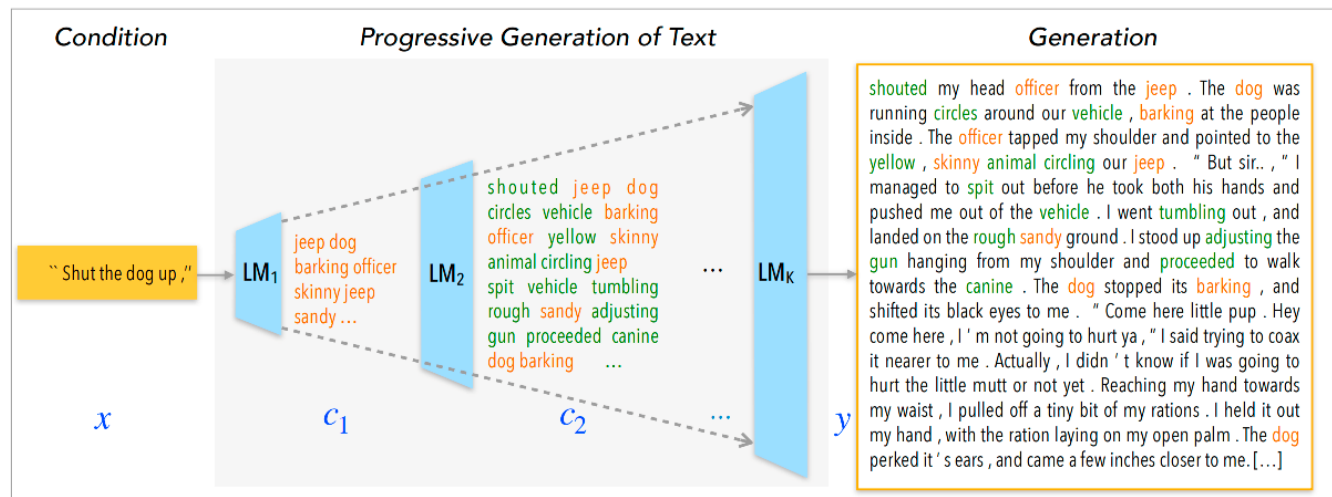
基于规划的长文本生成方法

- 这类方法将长文本生成过程分解成两个或多个阶段: 首先利用规划模型产生文本大纲, 然后基于文本大纲生成完整的文本, 如故事生成

两阶段生成方法

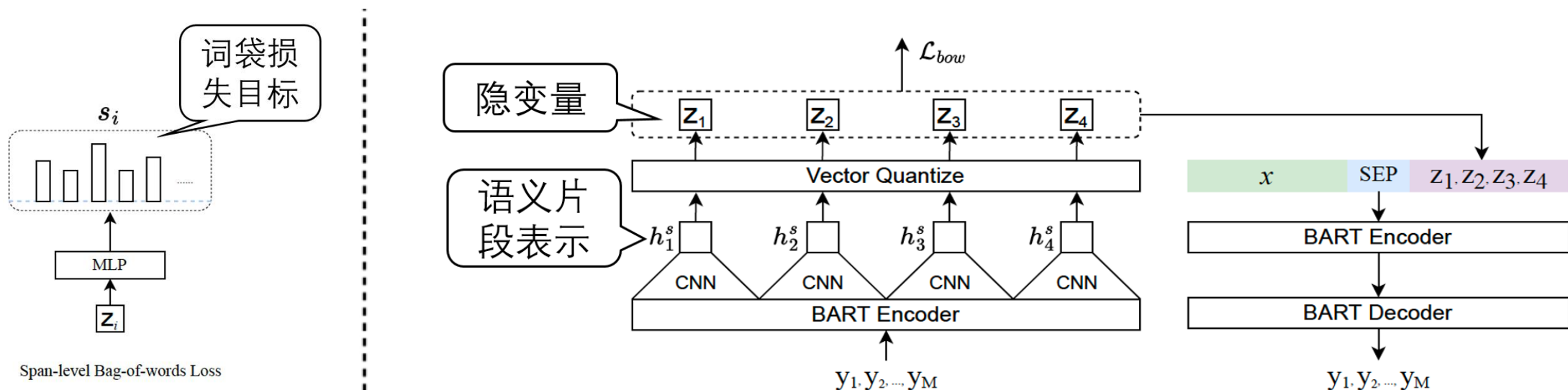


多阶段渐进式生成方法



该方法将长文本生成分解成更多的阶段, 首先生成一些重要的关键词, 然后通过多个阶段生成逐步添加细粒度的单词细化生成内容, 直到完成一个完整的故事

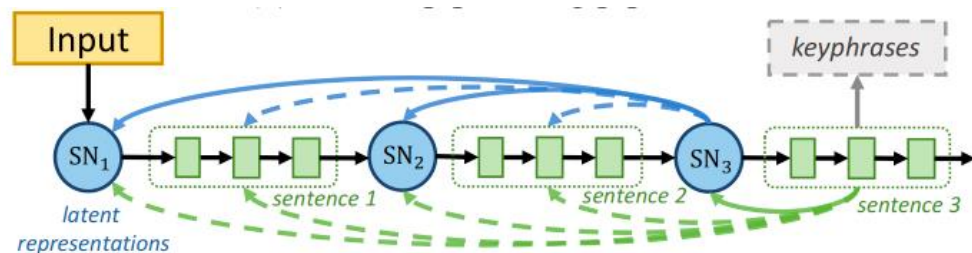
- 一篇自然的文本可以分割为多个连续的语义片段，每个片段围绕一个主话题展开
- 提出片段级词袋重构目标，使离散隐变量建模每个语义片段中的话题信息
- 利用话题感知的隐变量序列指导文本生成，使生成的内容与输入更相关且语义片段之间具有语义关联



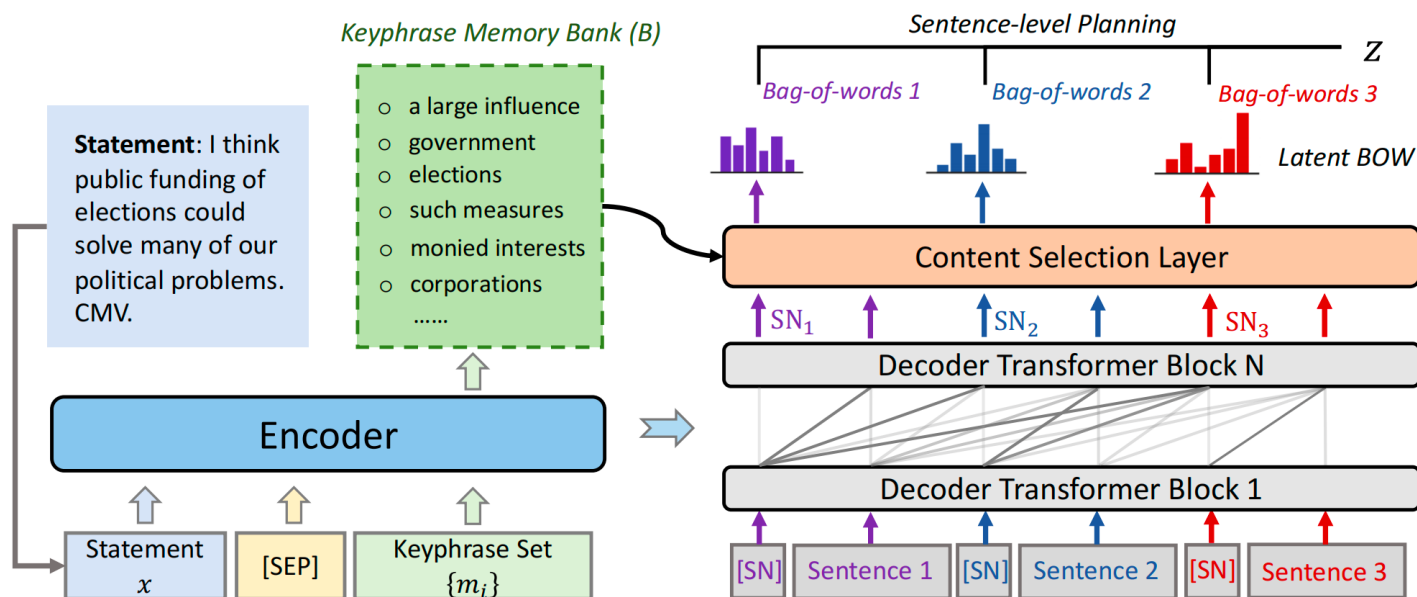
基于隐变量的长文本生成方法

- 两阶段生成方法中的问题
 - 规划模型和生成模型是分离的，存在错误累积问题
- 基于动态内容规划的方法
 - 将规划模型和生成模型联合在同一个模型中
 - 每个句子之前拼接一个特殊字符[SN]，提出句子级的词袋损失，使特殊字符[SN]学习待生成句的整体语义
 - 生成过程中，首先生成句子表示[SN]，随后[SN]指导生成整个句子
 - 生成时，基于注意力机制动态选择内容关键词

基于注意力机制的动态内容规划方法



实现动态内容规划的神经网络模型结构



基于记忆网络的长文本生成方法

- 基于transformer的预训练模型，处理长文本时面临以下问题：
 - 预训练模型支持有限长度的输入或输出文本，比如BART, GPT2一般最大支持1024个token，较难处理过长的文本
 - Transformer模型的计算量随序列长度n平方增长，即计算复杂度为 $O(n^2)$ ，处理长文本的计算量更大
- 改进方法
 - 包含多个段落的文章，采用逐段落生成的方法
 - 引入记忆单元（向量矩阵）记忆历史语义信息，从而维持生成文本语义上的前后连贯性

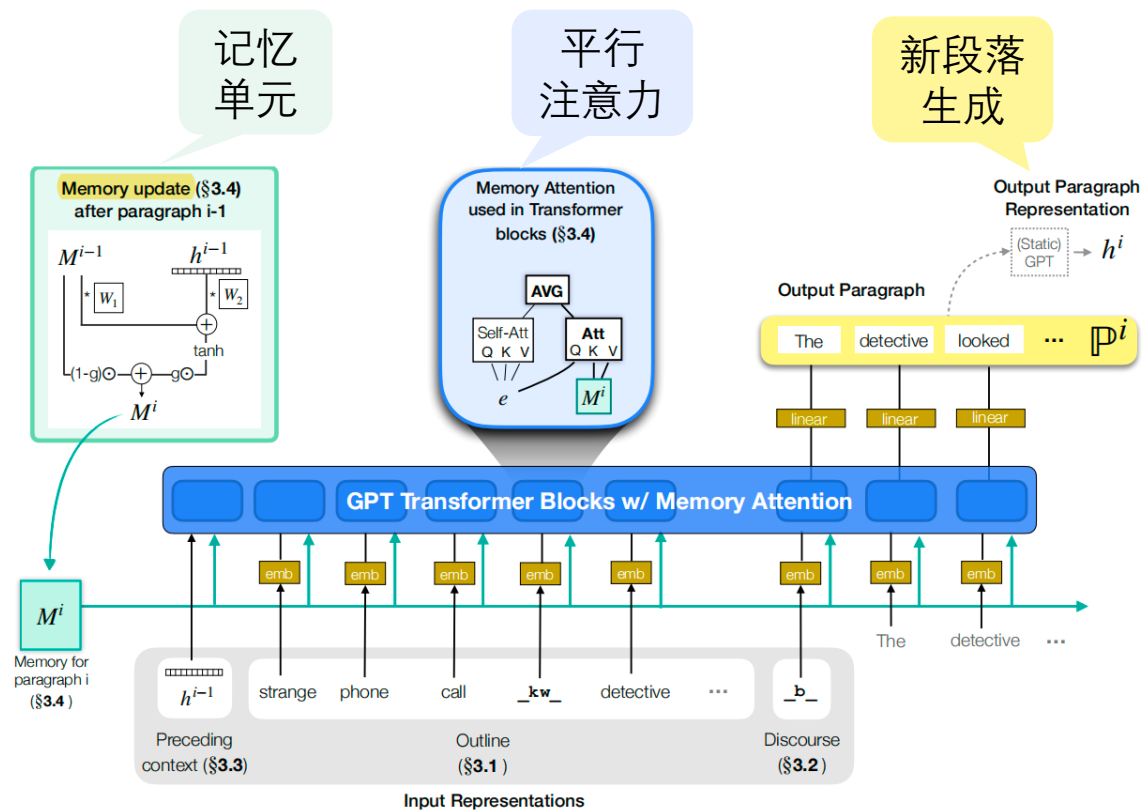
新段落生成和记忆单元更新方法

$$(\mathbb{P}^i, h^i, M^i) = \text{PM}(o, d^i, h^{i-1}, M^{i-1}) \quad (1)$$

$$\hat{M}_j^i = \tanh(W_1 M_j^{i-1} + W_2 h^{i-1}) \quad (2)$$

$$g_j^i = \text{sigm}(W_3 M_j^{i-1} + W_4 h^{i-1}) \quad (3)$$

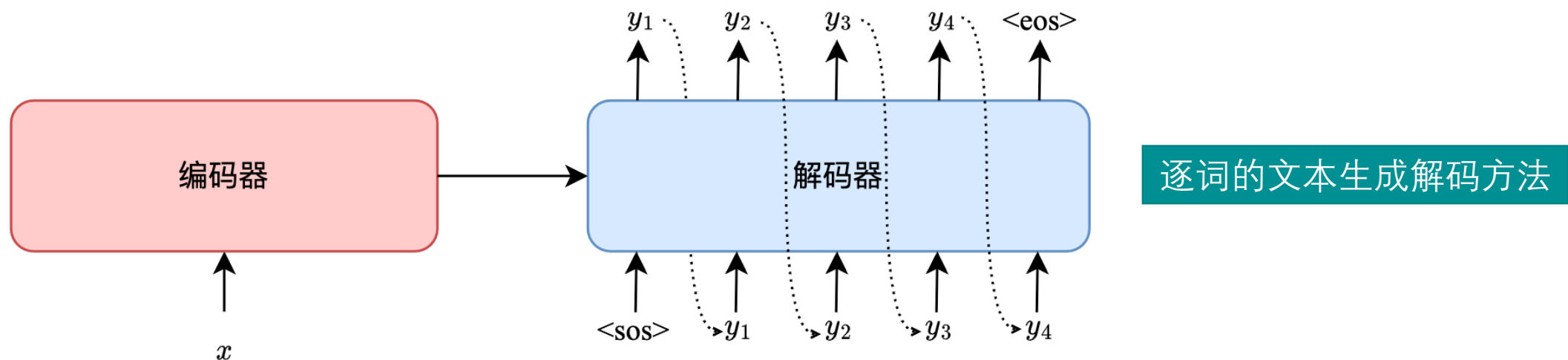
$$M_j^i = (1 - g_j^i) \odot M_j^{i-1} + g_j^i \odot \hat{M}_j^i \quad (4)$$



逐段落的长文本生成方法

- 01 文本生成基本方法与应用
- 02 文本生成中的可控方法研究
- 03 文本生成中融入常识、知识的方法研究
- 04 长文本生成方法研究
- 05 文本生成中的解码方法研究**

- 神经文本生成模型采用逐词解码方式，每一步预测一个词并作为下一步的输入

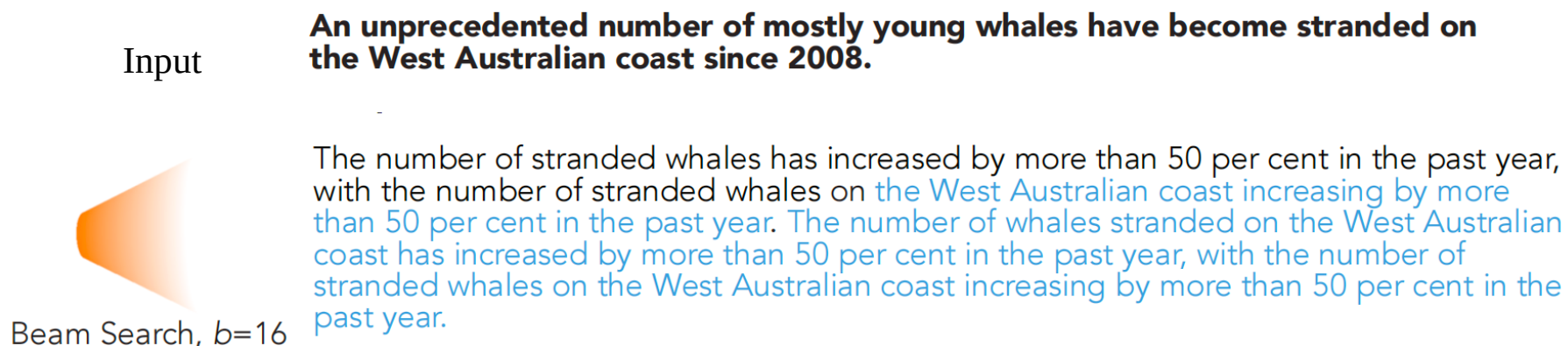


- 对于输入 x ，利用训练好的模型找到最佳生成文本 $\hat{y}$ 使生成概率达到最大的过程被称为解码，是一个搜索问题，搜索空间为 $|V|^T$ ， $|V|$ 是词表大小， T 是序列长度

$$\hat{y} = \operatorname{argmax}_y p(y|x)$$

- 但穷尽所有可能文本，时间、空间复杂度都非常高

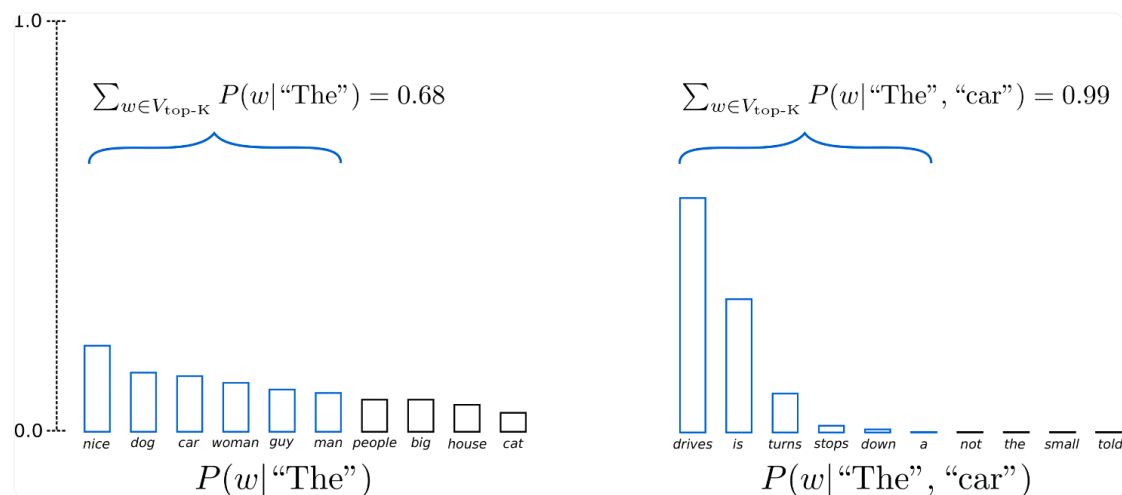
- 神经文本生成模型常用解码策略
 - Greedy Search: 每一步解码都选择最高概率的词作为目标词
 - Beam Search: 预设柱大小为 n , 每一步解码选出 n 个概率最高的作为目标词并同时只维持得分最高的 n 个候选序列, 剪枝掉其他序列
- 在开放式生成任务中, 如故事生成和文案生成, 输出的文本长度较长, 这些解码策略存在重复生成的问题, 包括词语级、短语级、句子级的重复, 导致多样性差、可读性差等问题



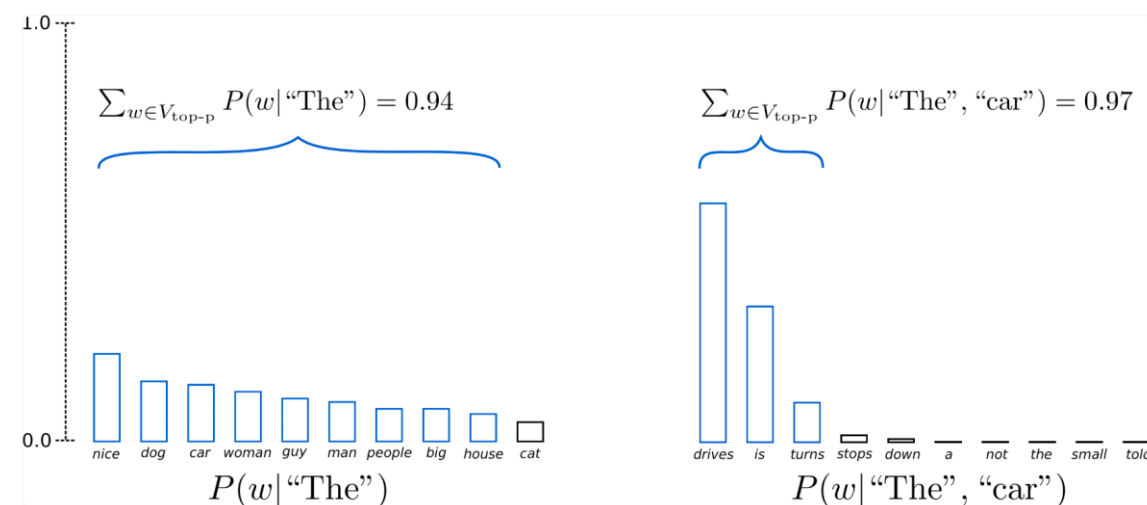
文本生成中的重复生成示例

基于采样方法改进重复生成问题

- Sampling decoding: 解码过程中引入随机性
 - Top-K sampling: 每一步选择概率最高的K个token构造采样空间，随后从中采样得到目标词
- Top-p sampling: 从概率最高的token开始构造采样空间，直到采样空间里所有token的概率加和超过阈值p，随后从中采样



Top-K构造采样空间



Top-p构造采样空间

- 基于最大似然估计训练模型的问题
 - 学习到的词向量表示具有各向异性分布（向量驻留在整个空间的一个狭窄子集中）

- 改进方法

- 在训练阶段，采用对比学习更好地进行表示学习

$$\mathcal{L}_{\text{CL}} = \frac{1}{|\mathbf{x}| \times (|\mathbf{x}| - 1)} \sum_{i=1}^{|\mathbf{x}|} \sum_{j=1, j \neq i}^{|\mathbf{x}|} \max\{0, \rho - s(h_{x_i}, h_{x_i}) + s(h_{x_i}, h_{x_j})\},$$

$$s(h_{x_i}, h_{x_j}) = \frac{h_{x_i}^\top h_{x_j}}{\|h_{x_i}\| \cdot \|h_{x_j}\|}.$$

$$\mathcal{L}_{\text{SimCTG}} = \mathcal{L}_{\text{MLE}} + \mathcal{L}_{\text{CL}},$$

- 在解码阶段，采用对比搜索，对同一句话中语义相似的token施加惩罚，降低重复生成，以生成多样和连贯的文本

$$x_t = \arg \max_{v \in V^{(k)}} \left\{ (1 - \alpha) \times \underbrace{p_\theta(v | \mathbf{x}_{<t})}_{\text{model confidence}} - \alpha \times \underbrace{(\max\{s(h_v, h_{x_j}) : 1 \leq j \leq t-1\})}_{\text{degeneration penalty}} \right\}$$

GPT学到的词语语义向量分析

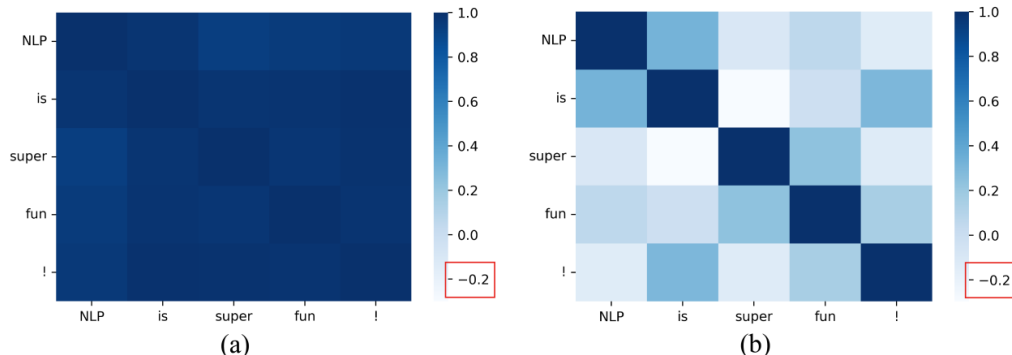
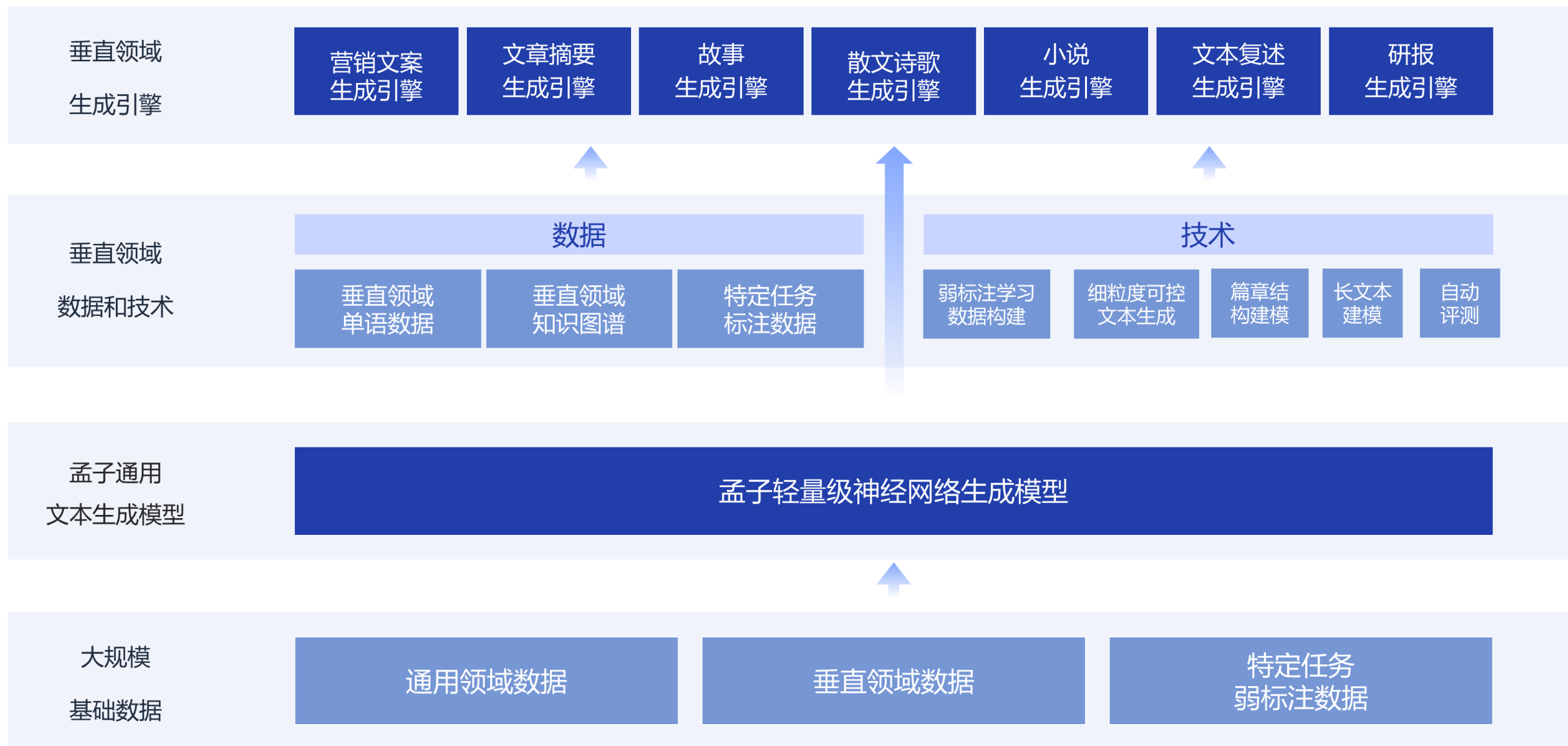


Figure 1: Token cosine similarity matrix of (a) GPT-2 and (b) SimCTG. (best viewed in color)

- 本报告介绍了目前常用的可控文本生成技术
 - 可控生成算法、常识和知识融入、长文本生成、文本生成解码
- 未来需要探索的方向
 - 生成文本的可控性目前集中在情感、关键词等方面，需要探索篇章结构可控、多样性可控、细粒度可控等
 - 融入常识、知识目前利用知识图谱的三元组的方法存在着知识获取难度高、使用武断等问题，需要有效改进
 - 长文本生成需增强模型对主题一致性、事实一致性、文章层次结构和前后逻辑等的学习，以及改进记忆网络的能力
 - 进一步改进多样化解码能力，单句级别甚至跨句级别
- 文本生成的评价体系、评测集、自动评测方法
- AI伦理：如何防止安全性、有害性的生成

澜舟公司部分文本生成项目介绍





多属性可控文本生成

通过灵活可定制的多属性输入形式，可输入如关键词、主题、情感、文本风格等控制属性，生成满足约束条件的文本，得到更加新颖和多样的文本。



多样化的文本生成形式

支持基于已有文本的续写，以及基于关键词、标题、表格数据的文本扩写，使用更加灵活便捷。



基于知识图谱的文本生成

用户可自定义知识图谱，有效提高生成文本内容的事实正确性。



长文本建模

支持千字以上的长文本生成，适用于多种写作场景，如汽车文案生成、小说创作等。



轻量化微调

在孟子预训练语言模型的基础上，构建轻量化微调模块，领域适应能力强，针对新任务可快速验证效果，模型训练成本低。



多任务联合建模

支持多任务参数共享，一个模型可支持多种场景，模型部署成本低。



内容和风格定制化

基于用户数据构建专属文本生成引擎。



文本自动评测系统

可以评测文本的事实正确性、逻辑连贯性等，并检测敏感内容、进行语法纠错，让生成文本的质量有保障。



多行业领域支持

支持营销、文学、金融、医疗、教育等众多行业。

澜舟文学辅助创作应用

基于关键词的内容生成



基于关键词生成 自定义模版生成 实体渲染 续写

输入内容仅限中文

选择文体风格：

☒ 都市 ☐ 玄幻 ☐ 修真

输入核心关键词（必填，文本将围绕关键词进行生成。以逗号分隔，逗号不区分中英文）

三月, 湖边, 翠绿, 垂柳, 飞鸟

输入辅助短语（选填，辅助生成文本中的细节描述。以逗号分隔，逗号不区分中英文）

柳树垂下绿丝绦, 掀开珠帘一角, 露出一泓清澈的湖水

随机示例

开始生成

输出结果：

珠帘一角掀开，露出一泓清澈的湖水。湖边垂柳依依，柳树垂下绿丝绦，飞鸟在翠绿的柳叶间穿梭。三月的湖边，绿意盎然，一片生机勃勃。

免责声明： 本页面提供的生成结果系模型自动生成，均不代表本平台立场

基于关键词生成 自定义模版生成 实体渲染 续写

输入内容仅限中文

选择文体风格：

☐ 都市 ☐ 玄幻 ☒ 修真

输入核心关键词（必填，文本将围绕关键词进行生成。以逗号分隔，逗号不区分中英文）

天边, 夕阳, 下坠, 娇红, 艳紫, 山峦

输入辅助短语（选填，辅助生成文本中的细节描述。以逗号分隔，逗号不区分中英文）

渲染出一种凄美的韵致, 顾盼着满天的云霞

随机示例

开始生成

输出结果：

夕阳西下，天边一片艳紫，顾盼着满天的云霞，渲染出一种凄美的韵致。娇红的山峦，随着夕阳的下坠，渐渐的暗淡了下来。

免责声明： 本页面提供的生成结果系模型自动生成，均不代表本平台立场

澜舟文学辅助写作免费体验地址：
<https://cognitive.langboat.com/product/literary-creation>

基于模板的内容生成

基于关键词生成

自定义模版生成

实体渲染

续写

 输入内容仅限中文

选择文体风格：

☐ 都市 ☒ 玄幻 ☐ 修真

输入自定义模板：（必填，用任意条数的下划线标示出待补全内容）

她有__的长发，__的眼眸，__的贝齿，__的樱唇。

随机示例

开始生成

输出结果：

她有着一头乌黑亮丽的长发，一双清澈明亮的眼眸，弯弯的柳眉，小巧的琼鼻，洁白无瑕的贝齿，娇艳欲滴的樱唇。

基于关键词生成

自定义模版生成

实体渲染

续写

 输入内容仅限中文

选择文体风格：

☐ 都市 ☐ 玄幻 ☒ 修真

输入自定义模板：（必填，用任意条数的下划线标示出待补全内容）

她是__，一双__，__，看着真是一位__。

随机示例

开始生成

输出结果：

她是典型的江南女子，一双水灵灵的大眼睛，身形苗条，看着真是一位美人。

免责声明：本页面提供的生成结果系模型自动生成，均不代表本平台立场

澜舟文学辅助写作免费体验地址：

<https://cognitive.langboat.com/product/literary-creation>

基于关键词生成 自定义模版生成 **实体渲染** 续写

① 输入内容仅限中文

选择文体风格：

☐ 都市 ☒ 玄幻 ☐ 修真

选择实体：

男性相貌 ^

女性相貌

男性相貌

武器法宝

山

夜晚

奇珍异兽

技能

战争

职业

门派势力

类人

输入辅助短语（选填，辅助生成文本中的细节描述。以逗号分隔，逗号不区分中英文）

浑身上下散发着冷峻

随机示例

开始生成

输出结果：

他身材挺拔，相貌英俊潇洒，浑身上下散发着冷峻，给人一种拒人于千里之外的感觉。

免责声明：本页面提供的生成结果系模型自动生成，均不代表本平台立场

基于关键词生成 自定义模版生成 **实体渲染** 续写

① 输入内容仅限中文

选择文体风格：

☐ 都市 ☐ 玄幻 ☒ 修真

选择实体：

武器法宝 v

输入核心关键词（必填，文本将围绕关键词进行生成。以逗号分隔，逗号不区分中英文）

长矛

输入辅助短语（选填，辅助生成文本中的细节描述。以逗号分隔，逗号不区分中英文）

带着呼啸的狂雷之声

随机示例

开始生成

输出结果：

长矛与巨斧碰撞，带着呼啸的狂雷之声，轰然炸响。这是何等可怕的力量？

免责声明：本页面提供的生成结果系模型自动生成，均不代表本平台立场

[基于关键词生成](#)[自定义模版生成](#)[实体渲染](#)[续写](#)

 输入内容仅限中文

选择文体类型：

☐ 文言文续写

☒ 现代文续写

输入故事的开头（必填）

只是在休息室里打了个盹儿，一睁眼，竟然穿成了古代目不识丁的乡下胖丫头。

35/200

随机示例

开始生成

输出结果：

只是在休息室里打了个盹儿，一睁眼，竟然穿成了古代目不识丁的乡下胖丫头。穿越了也就算了，还变成了这个身份。这让她很不适应。好在她有自知之明：自己什么也不懂！所以她没有去纠结这些小问题。而是开始打量这个世界-这是一个陌生的世界！她记得自己好像是在那个山洞中醒来，然后发现自己身上穿着古装，头上还有钗环玉佩什么的，于是就想着穿越到一个类似的地方。后来发现是穿越到了一个现代世界-南朝陈国，而这里就是她的出生地-陈都，而她的母亲和哥哥们都在这个时代。那么，既然穿越到了这个地方，那是不是说明自己是在这个世界生活一辈子？

免责声明：本页面提供的生成结果系模型自动生成，均不代表本平台立场

澜舟文学辅助写作免费体验地址：

<https://cognitive.langboat.com/product/literary-creation>

[基于关键词生成](#)[自定义模版生成](#)[实体渲染](#)[续写](#)

 输入内容仅限中文

选择频道：

☒ 男频 ☐ 女频

选择文体类型：

☐ 都市 ☐ 玄幻 ☐ 仙侠 ☐ 古风 ☒ 文言文

输入故事的开头（必填）

明有奇巧人曰王叔远，能以径寸之木，为宫室、器皿、人物，以至鸟兽、木石，罔不因势象形，各具情态。

48/256

随机示例

开始生成

输出结果：

明有奇巧人曰王叔远，能以径寸之木，为宫室、器皿、人物，以至鸟兽、木石，罔不因势象形，各具情态。此书所以称奇巧也。然天地之用，各有定形，各有定式。或如鬼神，或如草木丛生，其形状与人异；或如牛马驼骡，其形态亦与天殊；或如人身一身，犹如五脏之各有六腑。故天有五行、日月、风雷、雨雪，地有五物、日星丽焉；人物之生死荣枯寿夭，亦各有定数。

免责声明：本页面提供的生成结果系模型自动生成，均不代表本平台立场

澜舟文学辅助写作免费体验地址：

<https://cognitive.langboat.com/product/literary-creation>

澜舟营销文案生成应用

内容大盘

热点追踪

智能文案

视频创作

达人优选

发布管理

效果监测

内容电商

AI自动撰写

VS

人工撰写

平均每篇价格

1.5-3元

50-100元+

每篇产生的耗时

5秒以内

1小时起

风险管控

自动敏感词过滤

需要人工排查

经验转移

基于百万级历史优秀文章训练

更多需要人的经验

contentnote

化妆品领域[面膜]		
输入	标题	让你的肌肤重返18岁
	关键词	[姜汁, 美白, 面膜, 肌肤, 祛斑, 生活, 饮食, 护理]
	知识图谱	[肌肤问题, 困扰, 很多亚洲女性] [很多女性, 关心, 祛斑美白的问题]
孟子	现在困扰很多亚洲女性的肌肤问题之一就是黑色素的沉淀, 所以在日常生活中, 我们需要做到以下几点: 1、注意饮食, 多吃新鲜蔬菜水果, 少吃辛辣刺激性食物; 2、保持良好的作息時間, 不要熬夜; 3、坚持每天敷面膜, 每周敷2-3次; 4、多运动, 增强身体的新陈代谢; 5、合理的护肤, 选择适合自己肤质的产品; 6、适当的做一些面部按摩, 促进血液循环, 改善皮肤的状态; 7、调整好心态, 放轻松的心情, 做事情要有计划; 8、不过度的追求完美, 要学会享受当下的美好时光; 9、在平时的日子里尽量保证充足的睡眠和休息, 这样才能更好地保护自身的身心健康, 提高自我修复的能力; 10、每个人的体质都不一样, 如果你是敏感肌的话, 可以用温和的洗面奶进行清洁, 然后再用姜汁洗脸, 用生理盐水浸湿化妆棉, 敷在脸上15分钟左右, 洗掉之后再涂上一层保湿的面霜, 最后在用一個补水的精华液, 就ok了。另外, 还有一点很重要的, 那当然是要好好护理, 因为我发现现代人都是比较注重自律的人, 在这个过程中我也会更加的关注一下我的生物钟, 比如说我是属于那种经常会失眠的类型, 而且我还是一名职场白领, 平常的工作压力也是非常大的。在我看来, 只要我能够调节好我平日里的一日三餐, 也就不担心会长斑啦。	
GPT	肌肤问题一直是困扰着很多亚洲女性的,所以很多女性都很关心祛斑美白的问题,那么如何美白肌肤呢?今天我就来给大家分享一下我的美白小妙招,让你的肌肤重回18岁!1.饮食方面:饮食清淡,少吃辛辣刺激性食物,多吃富含维生素C的食物,如:胡萝卜、西兰花、番茄、西红柿等。2.日常护理:注意防晒,做好防晒工作。3.日常护肤:坚持使用美白产品,比如美白精华、美白面膜、姜汁面膜等等。4.日常保养:保持良好的心情,保持愉快的心情,不要熬夜。	

数码产品领域[电脑包]		
输入	标题	电脑包推荐
	关键词	[电脑, 保温杯, 电脑包, 包包, 平时, 袋子, 入手, 背着]
孟子	最近入手了一款超级好用的电脑包, 就是这款包包啦, 我平时工作比较忙, 所以经常需要出差, 但是又不想带太多的东西, 这个包的容量还是很大的, 可以放下我的ipad, 还有一个保温杯和一些日常用品, 而且它的袋子也是非常的好看, 是那种很有质感的牛皮材质, 摸起来手感也很好, 很高级的感觉。我买的是黑色的那款, 因为我是学生党嘛, 也不太喜欢太花里胡哨的装饰品。这样一来就不会显得很突兀, 反而会让整体的搭配更加的和谐。它里面的内部空间也挺充足的呢, 有两个小的收纳袋, 放笔记本和ipad是完全没问题的哦, 平常出门的时候我都会带上它, 不仅能装下笔和电子产品还能放些小零食, 比如说我们的口红啊, 钥匙啊等等, 真的太方便了。最重要的一点是, 它还送了两根挂绳, 背着它去哪里都不用担心会掉下来哦。而我最喜爱的还要数它那个可拆卸的小布袋了, 里边有很多小袋袋的设计, 像我这种手残党也能轻松驾驭, 出去旅游或者是去公司办公都很适合背它哦! 最后我想说一下, 如果你也想拥有同款的话, 建议你们也去试一试, 毕竟这么便宜的价格, 性价比也太高了吧! 我现在已经成为了它们家的忠实粉丝了呢!	
GPT	最近入手了一款电脑包,容量很大,平时可以装下电脑、手机、充电宝、保温杯、口红、粉饼、钥匙、纸巾等等。这款包包的材质是牛皮的,摸起来手感很好,而且很轻,背着也不会觉得重。我平时上班背电脑比较多,所以这款包的容量也很大。这个包包是双面的,可以放电脑,也可以放手机,还可以放充电宝,还有两个袋子,一个袋子装着电脑和手机。平时背着它去上班,也很方便。	

孟子模型中不同输入的性能评测

孟子基础模型	B-1↑	B-2↑	B-4↑	R-1↑	R-2↑	Dist-1↑	Dist-2↑	Conciseness↑	Controllability↑	PPL↓
标准答案	100	100	100	100	100	79.7	95.2	45.4	100	43.0
输入标题	20.5	7.0	4.3	23.2	7.9	77.2	93.7	32.1	75.0	30.1
输入标题+关键词	31.2	16.1	11.9	36.3	18.7	75.5	92.7	32.5	80.3	28.3
输入标题+三元组	28.2	15.3	13.2	32.2	17.4	77.1	93.3	34.7	81.6	27.9
输入标题+关键词+三元组	37.7	23.4	19.5	43.6	27.0	77.4	93.7	35.6	82.1	27.7

不同预训练模型的性能评测

模型	B-1↑	B-2↑	B-4↑	R-1↑	R-2↑	Dist-1↑	Dist-2↑	Conciseness↑	Controllability↑	PPL↓
孟子基础模型	37.7	23.4	19.5	43.6	27.0	77.4	93.7	35.6	82.1	27.7
孟子可控文本生成模型	39.3	24.6	20.5	44.7	27.7	78.1	94.4	36.9	86.0	26.8
GPT	37.1	23.0	19.2	43.3	27.1	78.0	94.2	36.8	56.0	26.1

这款AI产品生成的内容，竟然比真人更专业？！

2021年9月，数说故事与北京澜舟科技正式签订合作协议，双方在NLP领域展开深度合作。数说故事旗下专注于内容智能的子公司——横琴容徽，在澜舟科技的技术支持下，打造了一款**自动化写作产品——contentnote智能文案**，目前仅应用于营销文案的自动生成，12.30正式对外发布

仅仅三个步骤（选择模版-输入关键词-生成结果），营销人员便可通过contentnote生成高质量、原创的营销文案，为营销带来价值：

- 减轻创作者的工作压力，提高工作效率
- 让职场新人也能成为文案高手



contentnote容徽



营销文案生成产品示例

请选择写作模板

全部 营销文案 非营销文案

 产品营销 unlimited narrative methods and mouth-to-mouth, for product selling points and advantages for promotion	 好物推荐 sharing the way of personal experience, combining personal feelings and stories, recommending a product	 使用教程分享 as a deep user, sharing the experience and experience, while promoting the product	 产品测评 after real experience, product review, sharing product information	 开箱分享 sharing the unboxing process
 专业科普 incorporating professional popularization information, and introducing the product	 带货转化 coming out	 品牌形象宣传 coming out	 品牌故事 coming out	 个人体验/剧情 coming out

标题 ① 牙医推荐的牙膏，你值得拥有！

* 品类 ① 牙膏

品牌名 舒客 产品名 例：小棕瓶

* 利益点与其他 ① 颜色, 美白, 益生菌, 牙齿敏感, 抑菌

没有想法？点击 [生成推荐词](#) ①

白色 牙齿敏感 蛀牙 小苏打 绵密 舒适 家用 可爱 口腔溃疡 懒癌患者 妈妈 无氟 补水 3D 柔软 清爽 氨基酸 母婴 女性 温和 牙黄 含氟 夏季 一次性 抑菌 硅胶 软毛 益生菌 清新口气 口臭 按压式 白 清凉 水润保湿 薄荷 保护牙齿健康 细腻 冬天 免打孔 电动 出门 海盐 宝妈 牙龈出血 女儿 口香糖 黄牙

科技 痛点 功效 场景 2/2

生成条数: 7

清空条件

生成结果 历史记录

5 牙医推荐的牙膏，你值得拥有！

复制

舒客益生菌美白牙膏，这是一款可以让你牙齿变白的产品，颜色是透明的，看起来就很干净，而且它还有抑菌的功效，对于我这种牙龈敏感的人来说，真的是太友好了，用了一段时间之后，明显的感觉到自己的口腔健康了很多，也没有那么容易出现牙黄的情况了。它的外包装也很好看，是那种简约大气的设计，里面的膏体是淡淡的粉色，刷牙的时候不会有刺激感，很温和，味道也是清新的薄荷味，每次刷完牙嘴巴里都是香香的。坚持使用一个月，你就会发现你的牙变得白白净净的啦，牙垢也会变少，口气也变好啦。这款牙刷的刷毛也比较柔软，不伤害到牙釉质，还能够起到一定的清洁作用。我现在每天早晚都会刷一刷，效果还是很不错的哦，我用的这个是蓝色款，它有一点点的蓝白渐变，但是不是特别的夸张，就是我们平时所看到的那个蓝，其实是有点偏黄色的一种，如果你想要更加白一些的话，建议你选择白色那款哦。而我最喜欢的还要数它们家的另外两款颜值超级高的漱口水啦！这两个都很有特点，都有两种不同的香型，分别是玫瑰味和茉莉花味。每一次

Chinese

Brand name	面霜
Key words	抗老,平滑,丰盈,舒适,肌肤,肤感,,柔润舒适，胶原蛋白生成
Marketing copywriting	雅诗兰黛全新白胶原霜添加了日本紫珠果精粹，能够帮助肌肤抗氧化，同时促进胶原蛋白生成，有效抗老。质地轻盈，肤感柔润舒适，吸收快，不粘腻，使用后皮肤平滑细腻有光泽。搭配按摩手法，可以更好的改善细纹、松弛等问题，令肌肉更加紧致，丰盈饱满。用了一段时间后，明显感觉到脸部变得紧实有弹性了，而且肤色也提亮了不少，整个人看起来年轻了好几岁呢！

English

Brand name	Cream
Key words	Anti-aging, smooth, plump, comfortable, skin, skin feel, soft and comfortable, collagen production
Marketing copywriting	Estee Lauder's new White Collagen Cream is enriched with the essence of Japanese purple bead fruit, which can help the skin to antioxidants, while promoting collagen production and effective anti-aging. The texture is light, the skin feels soft and comfortable, absorbs quickly, is not sticky, the skin is smooth, delicate and shiny after use. Combined with massage techniques, you can better improve fine lines, sagging and other problems, so that the muscles are more firm, plump and full. After using it for a while, it is obvious that my face has become firm and elastic, and my complexion has brightened a lot, and I look a few years younger!

Japanese

Brand name	クリーム
Key words	アンチエイジング、スモース、ふっくら、心地よさ、肌、肌感、、しなやかで快適、コラーゲン生成
Marketing copywriting	エスティローダーの新しいホワイトコラーゲンクリームは、抗酸化作用を促進し、コラーゲンの生成を促進し、アンチエイジングに効果的です。軽いテクスチャで、肌が柔らかく快適で、吸収が速く、粘着性がなく、使用後の肌は滑らかで繊細で光沢があります。マッサージ手法を組み合わせることで、しわやたるみなどの問題をより良く改善し、筋肉をより引き締め、より豊かにすることができます。しばらく使用した後、明らかに顔が締め付けられ、弾力性があり、肌の色も明るくなり、全体的に何年も若く見えます！

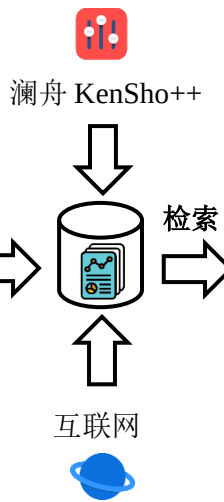
Portuguese

Brand name	creme
Key words	Anti-envelhecimento, suave, gordo, confortável, pele, sensação de pele, macio e confortável, produção de colágeno
Marketing copywriting	O novo creme de colágeno branco da Estee Lauder é enriquecido com essências japonesas que ajudam a pele a combater a oxidação e, ao mesmo tempo, estimulam a produção de colágeno para combater o envelhecimento. Sua textura é leve, suave e confortável, absorve rapidamente, não é pegajoso, após a aplicação, a pele é lisa, delicada e brilhante. Com a técnica de massagem, você pode melhorar as linhas finas, relaxamento e outros problemas, tornando os músculos mais firmes e cheios. Depois de usar por um período de tempo, senti claramente que o rosto tornou-se firme e elástico, e a cor da pele também brilhou muito, toda a pessoa parece mais jovem de alguns anos!

澜舟智能研报生成应用

智能研报

半导体
医疗服务
光伏设备
珠宝首饰
新能源汽车
通信服务
玻璃玻纤
消费电子
元宇宙



汽车行业周观点：1月汽车零售小幅下降，新能源汽车需求不悲观

新能源汽车行业周报：2030年新能源车与清洁能源汽车渗透率目标40%

公司首次覆盖报告：拐点将至，新能源汽车车载电源龙头扬帆起航

汽车行业深度报告：从电动化龙头看我国新能源汽车的进化之路

整合

问答
FAQ

Q: 某公司采用了怎样的市场策略应对新冠疫情的影响?
A: 适时开发..... 通过.....进行市场预热, 推出.....等手段增加现金流.....

事件
抽取

时间: XXXX 年 XX 月 XX 日
融资方: XX XX
事件: XX XX 融资余额 XX 亿元, 创近一年新高

摘要
生成

我们从政策、供给、需求三个维度判断 XXX 迎来爆发...
政策上..... (信息来源: XXXX)
供给上..... (信息来源: XXXX)

舆情
分析

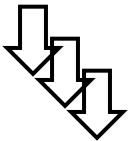
研报来源 1: 情感指数 -57, 关键词, 正负情感句高亮
研报来源 2: 情感指数 0, 关键词, 正负情感句高亮
研报来源 3: 情感指数 20, 关键词, 正负情感句高亮

研报
生成

标题 → 大纲生成 → 段落生成

输入：××公司ESG报告

大纲
生成



一、报告前言

（一）报告规范

报告时间范围：本报告为××公司第*份社会责任报告，报告时间范围为××年×月×日至×月×日，部分内容超出上述范围。

.....

（二）高管致辞

一年来，××公司的可持续发展工作得到了国家、社会和员工的支持和认可。××公司服务三农，保障三夏，为农业增产增收做贡献和赞助健康快车，为贫困地区白内障患者送去光明两项社会实践活动，入选××年国务院、国有资产管理委员会评选的‘中央企业优秀社会责任实践’；

.....

段落
生成



报告前言

- 报告规范
- 高管致辞
- 责任聚焦
- 企业简介

责任管理

- 企业使命、愿景、价值观
- 战略
- 组织
- 制度
- 文化
- 参与

市场绩效

- 股东责任
- 客户责任
- 伙伴责任

社会绩效

- 政府责任
- 员工责任
- 安全生产
- 社区责任

环境绩效

- 绿色管理
- 绿色生产
- 绿色运营

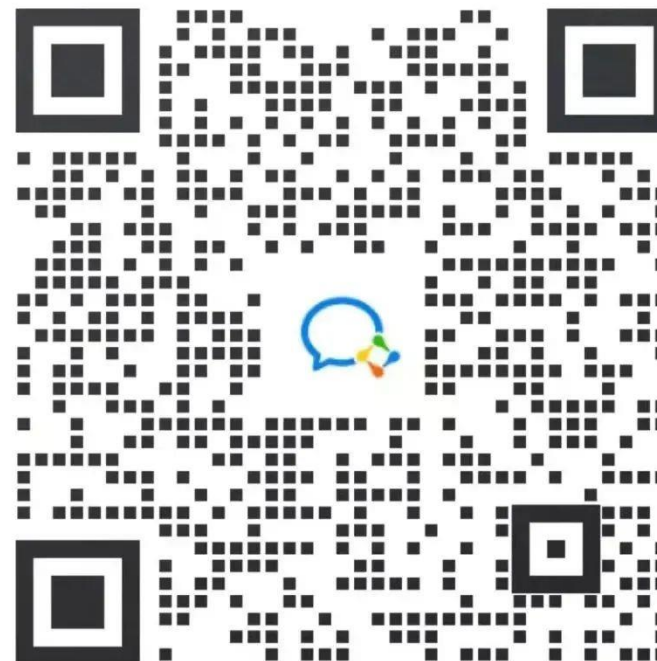
报告后记

- 未来计划
- 关键绩效表
- 企业荣誉表
- 报告评价
- 参考索引
- 意见反馈

澜舟熊猫小说家微信小程序应用

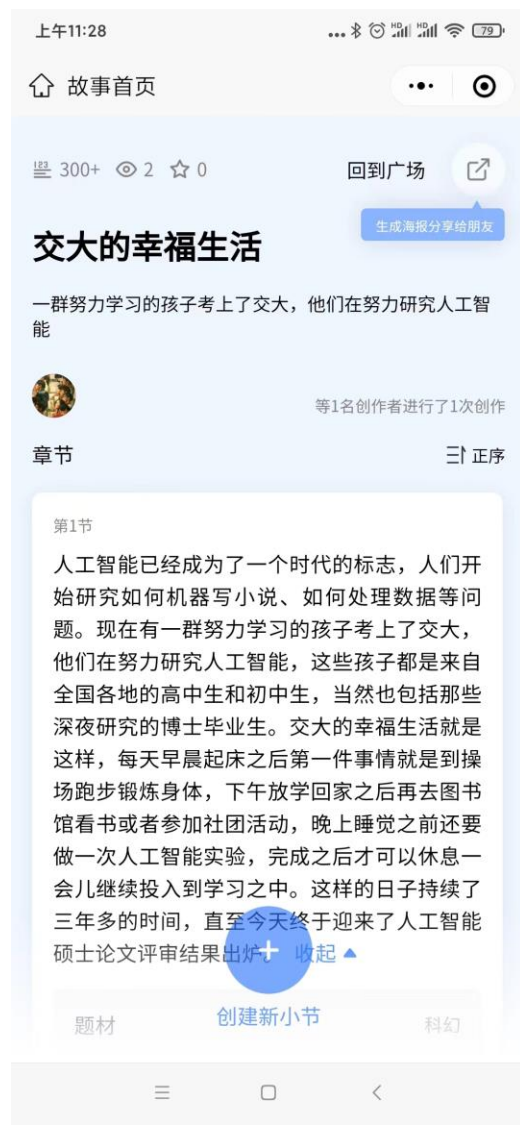


“澜舟熊猫小说家”微信小程序



“澜舟熊猫小说家”用户交流群

澜舟熊猫小说家应用



澜舟专业论文助写应用

基于作者输入的关键词进行扩写，生成包含用户期望信息的推荐例句

In this work, we proposed a NMT model with fast generation

Generate by these keywords

在此处输入生成时所包含的细节。最大支持100个字符

required

OK

Last configuration

transformer × decoder × sota × results show ×

Generate

in this work, we proposed a NMT model with fast generation
that uses the transformer seq2seq model with a one-to-one sota left-to-right decoder. Our results show that our model achieves better results in this work as compared to the autoregressive generation models like the baselines that have been used for NMT.

✓ | ↺ | ↻ | ✕

NLP style ▼

基于上文进行续写，生成推荐例句



in this work, we proposed a NMT model with fast generation|
that generates translation in real time and can be applied to real applications such as simultaneous interpretation. Our model is based on a Transformer model and uses a modified decoding procedure to speed up the generation process at every step.

✓ | ↺ | ✕

generate again
Alt+J



澜舟文图生成应用



作品信息

作品描述：长城、秋天、枫叶

关键词：充满细节的

下载



作品信息

作品描述：山坡、绿色、大树、小鸟

关键词：充满细节的, 高清画质

下载

Thank you!

北京澜舟科技有限公司

www.langboat.com

bd@langboat.com



关注“澜舟科技”公众号了解更多资讯



加入孟子开源社区探索NLP前沿技术

欢迎联系实习、工作和合作