

自然语言生成的方法及应用研究报告

李凯华
天津大学
kaihua_li@tju.edu.cn

摘要:自然语言生成是自然语言理解的一个重要研究领域，自然语言生成的研究目标是使机器能够生成人类能够理解的自然语言。自然语言生成在机器翻译、对话系统、报告生成、文章摘要、视频总结等多方面有广泛的应用场景。最早研究者们基于管道架构的模版技术、数学统计方法等实现自然语言生成，但相关方法高度依赖人工设计，并且在生成质量与复杂性上存在局限性，随着神经网络研究深入，研究者们开始利用神经网络构建端到端的自然语言生成，基于 Transformer 架构语言大模型成为热门的研究方向，其中自然语言生成中的幻觉现象也引起了研究者的关注。自然语言生成模型的评估方法也是研究难点之一。

1 引言

自然语言生成是自然语言理解的一个重要研究领域，自然语言生成的研究目标是使机器能够生成人类能够理解的自然语言。Reiter and Dale[1] 将自然语言生成定义为人工智能和计算语言学的一个分支领域，研究如何构建计算机系统，使其能够从某种非语言信息表征中生成可理解的英语或其他人类语言文本。

随着自然语言生成研究的深入，自然语言生成得到了广泛的应用，用于构建对话系统，使其能够自动生成自然、流畅的对话；用于机器翻译，能够生成更准确，更符合本地化用语习惯的译文；用于生成结构化的报告、摘要或分析，以便用户更好地理解数据和信息；用于生成文章摘要、视频总结，便于快速阅读和把握文章或视频内容。

传统的自然语言生成技术主要包括基于规则或模版的自然语言生成系统，这种方法使用预定义的规则和模板，根据输入数据生成文本，这通常用于生成结构化的文本，如报告或通知。其缺点是规则或模版的依赖人工设计，规则和模版针对特定领域设计导致泛用性较差，且其生成文本的复杂性和语义质量具有局限性。

Bengio[2] 在数学统计方法 n-gram 的基础上，应用神经网络技术并获得了显著的提高，随着进一步的研究深入，基于神经网络建模的方法在自然语言生成中得到了广泛的研究，神经网络技术相比规则或模版方法，省却了人工设计的昂贵步骤，依赖于丰富的训练语料，其生成文本具有形式丰富，语义准确等优越性，但是神经网络本身的训练代价和不可解释性是该技术的缺点。

2 自然语言生成方法

2.1 传统生成方法

2.1.1 模版生成方法

模版方法是指研究者根据人工设计出模版，在模版中存在可变量，根据实际的生成场景改变这些变量从而得到生成的模版。如

(今日)是(星期五)

，在该模版中“今日”与“星期五”是可变量。

模版生成方法具有可确定性、计算开销小等优点，但是其生成文本的形式缺乏变化具有很大的局限性。

Mcroy[3] 等人提出通过添加声明式的控制表达式和属性语法机制以增强传统模版方法，克服其文本生成形式单一的缺陷。

2.2 神经网络生成

2.2.1 RNN

循环神经网络 (Recurrent Neural Network, RNN) 借助内部的循环结构，使得序列数据先前的信息能够在网络中进行传递，能够很好地处理自然语言生成等序列任务。最早 Jodan[4] 设计了一种神经网络结构用于存储输出，并作为下一次输入的隐藏状态，其结构如图 1 所示。

Jordan 所设计的隐藏单元存储输出层的输出，Elman[5] 简化了 Jodan 的隐藏元结构，隐藏单元直接存储隐藏层的输出。现所使用的 RNN 一般指 Elman 所设计的循环网络结构。

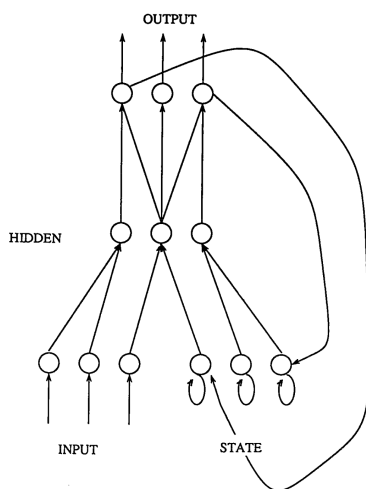


图 1: Jadon 设计的循环结构[4]

RNN 能够很好的学习到序列训练数据中的前后依赖关系, 在自然语言生成中有着优异的表现。但是 RNN 在训练过程中存在梯度消失和梯度爆炸的问题。

2.2.2 LSTM

长短期记忆神经网络 (Long-Short Term Memory, LSTM) [6] 通过引入门控单元解决 RNN 在反向传播过程中梯度消失和梯度爆炸的问题。

2.2.3 Seq2Seq/Encoder-Decoder

Cho[7] 提出了一种由两个 RNN 组成的 Encoder-Decoder 架构, 其中一个 RNN 将输入序列编码为固定长度的向量表示, 而另一个 RNN 将该向量表示解码为输出序列。Encoder-Decoder 结构如图 2 所示。

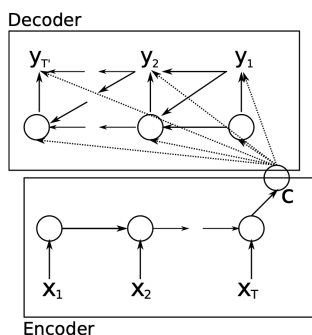


图 2: Encoder-Decoder 结构[7]

Sutskeve[8] 提出 Seq2Seq 模型, 利用 LSTM 实现端到端地输入序列到输出序列的映射。

Seq2Seq 与 Encoder-Decoder 架构解决了输入序列和目标序列不能固定长度编码的问题。此外 Seq2Seq 文章中还指出通过翻转源输入序列能够显著地提高模型表现。

2.2.4 Attention

Bahdanau[9] 推测 Encoder-Decoder/Seq2Seq 架构的模型表现性能瓶颈受中间结果固定长度向量的长

度影响, 并提出通过模型搜索目标序列和源序列之间的对应关系, 这种技术被称之为注意力机制(Attention Mechanism), 其结构如图 3 所示。

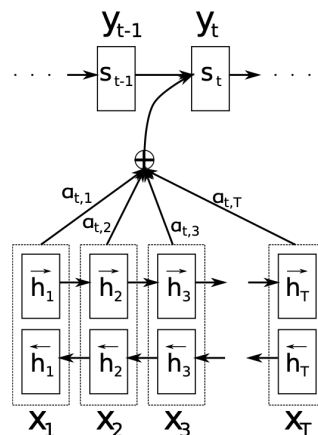


图 3: 注意力机制结构[9]

注意力机制的关键在于改变 Encoder-Decoder 架构中的解码器部分, 提出将原架构中的固定向量改变为 Context 向量, 即

$$p(y_i|y_1, \dots, y_{i-1}, \mathbf{x}) = g(y_{i-1}, s_i, c_i)$$

其中 c_i 是由输入序列按权重计算获得, 计算公式为

$$c_i = \sum_j^T \alpha_{ij} h_j$$

原始的 Encoder-Decoder 架构无区分地将原序列信息通过中间固定向量带入到解码器中, 而注意力机制模拟了人类的注意力特点, 能够观察到原序列中影响权重大的部分, 通过 α 权重参数控制原序列中不同单词在解码阶段的影响程度, 从而使模型达到更好的性能表现。

2.2.5 Transformer

Encoder-Decoder/Seq2Seq 架构中由于使用 RNN 或 LSTM 进行编码解码工作, 而 RNN 或 LSTM 为序列串行计算, 因此在计算过程中, 难以利用 GPU 进行加速计算。Vaswani[10] 提出 Transformer 架构, 使用 Self-Attention 机制完全代替了 RNN/LSTM 的编解码工作, 不仅能利用并行计算大大加速推理速度, 同时 Transformer 模型还有更好的质量表现。Transformer 由多个 Transformer Block 重复组合构成, 其结构如图 4 所示。

Transformer Block 由多头注意力层与前馈神经网络构成, 并且每层都结合残差链接和规范化操作。

研究者在 Transformer 结构上进一步深入研究, 发现 Transformer 在自然语言生成领域表现出优异的表现。

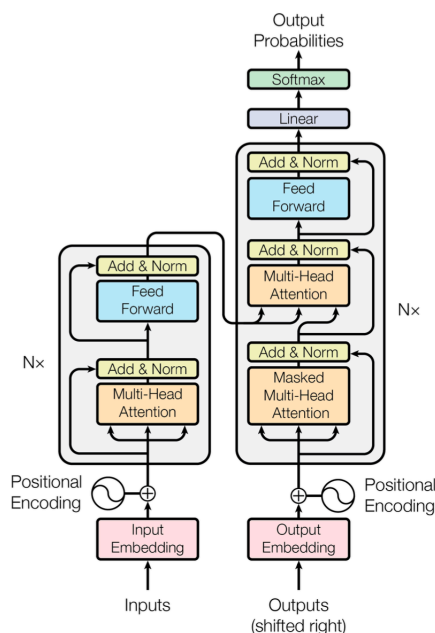


图 4: Transformer 结构

OpenAI Radford[11] 提出通过生成式预训练和个性化微调技术，基于 Transformer 架构提出 ChatGPT 系列对话系统，在自然语言生成方面获得显著成果，得到了广泛的关注。

3 自然语言生成应用

3.1 机器翻译

机器翻译的目的是使机器能够学会不同语言之间的映射关系，将一种形式的语言翻译成另一种形式的语言。借助机器翻译能够实现跨语言的沟通 and 理解，使不同语言的使用者之间的交流更加便捷。

Stahlberg[12] 调研了机器翻译的发展历史，指出统计方法的机器翻译主要应用于特定的翻译领域，目前由神经网络构建的翻译系统已经逐渐代替统计方法的机器翻译，成为主流研究方向。

早年 Bengio[2] 利用前馈神经网络增强机器翻译，随着神经网络在机器翻译中的进一步发展，Schwenk[13] 使用前馈神经网络直接给短语组进行评分，Cho[7] 使用 RNN Encoder-Decoder 架构进行机器翻译。所有这些方法都应用了神经网络作为传统统计机器翻译系统的组成部分。因此，他们保留了对数线性模型组合，只替换了传统架构中的部分结构[12]。

神经机器翻译 (Neural Machine Translation, NMT) 通过使用单个大型神经网络将源句子直接转换为目标句子来改变传统架构中的分离组件，神经机器翻译的出现成为机器翻译的里程碑[12]。Sutskever[8] 提出的 Seq2Seq 工作是端到端神经机器翻译的代表工作之一，该工作使用单一整体的神经网络实现原始语言到目标语言的翻译工作，其将 BLEU 分数由 33.3 提升至 36.5，展现出优异的表现。

3.2 对话系统

对话系统是指使机器能够根据人类的输入内容，生成对应的回应，以模拟人与人之间自然语言交流。对话系统基于规则引擎或人工智能技术实现，低智能的对话系统只能根据特定规则完成特定场景的对话，而基于人工智能的高智能对话系统能够根据各种复杂的对话场景给出对应的合适回复。

对话系统在生活中有着广泛的应用，也是自然语言理解中的热门研究方向，其涉及多种自然语言理解技术，包括卷积神经网络、循环神经网络、注意力机制及 Transformer 架构、强化神经网络与对抗神经网络等各种技术。

对话系统按架构又可以分为面向任务的对话系统与开放领域对话系统，开放领域对话系统又称为聊天对话系统。

面向任务的对话系统旨在解决特定领域的任务，与聊天对话系统不同，面向任务的对话系统需要回复准确的答案已解决用户的任务需求，如外卖机器客服、网购机器客服等。面向任务的对话系统由自然语言理解、对话状态跟踪、策略学习、自然语言生成等多个部分构成。Wei[14] 提出一套面向自动诊断的对话系统，近年有研究者尝试将对话系统定义为纯粹的自然语言生成任务，以利用 ChatGPT 的强大能力[15]。

开放领域对话系统或聊天对话系统，旨在处理多样性和开放性的用户输入，而不限定于特定任务。开放领域对话系统可以分为生成式系统、检索式系统与混合式系统。生成式系统是指以 Seq2Seq[8] 架构将用户的输入内容序列映射为回复序列的模型及系统，检索式系统则是从预设的回复集合中寻找合适的回复答案。混合式系统则是生成式系统与检索式系统的结合。

3.3 报告生成

报告生成式自然语言生成的重要应用领域，报告生成是指通过自动化的方式生成书面形式的文本报告，涵盖了各种领域，如金融领域股票报告、医学领域诊断报告、传播领域新闻报告等。

Boag[16] 的工作介绍了通过计算机视觉与自然语言生成结合实现 X 射线诊断报告生成系统。Varges[17] 提出的 SemScribe 基于医学领域的知识构建，能够生成针对心脏病检测医生诊断包括。

3.4 文章摘要

文章摘要旨在从文本中提取关键信息并以概括性的总结呈现给用户，方便用户从海量资料中快速掌握信息。

文章摘要一般有抽取式和生成式两种实现路线，抽取式是指从文章中抽取关键词、关键句以构建概括性总结，而生成式则是一种端到端的实现方式。Rush[18] 提出利用基于 Attention 机制的完全数据

驱动的神经网络模型实现生成式的文章摘要。Nallapati[19] 提出 Summarunner, 一种基于 RNN 循环神经网络的文章摘要模型。

4 自然语言生成幻觉

基于 Transformer 架构与预训练技术的大模型在自然语言生成方面表现出惊人的能力, 在机器翻译、对话系统等自然语言生成研究领域中有广泛的应用, 但是大模型自有的幻觉现象, 使得在面向任务、高度依赖确定性的领域难以广泛应用, 如医药相关领域、专业知识领域。Ji[20] 对自然语言生成大模型中的幻觉现象进行了广泛的调研。

自然语言生成中出现幻觉现象, Ji[20] 将幻觉定义为生成内容相对于原内容式无意义的或不可信的。自然语言生成中出现的幻觉与人类的心理现象有着相似的表现, 即感觉到不存在的事物, 在神经网络中表现为原内容与输出内容之间缺乏可靠关联。

Ji[20] 指出幻觉现象的出现来自于两方面, 一方面主要是数据本身存在源-引用分歧(source-reference divergence), 另一方面是神经网络模型的训练和建模。

幻觉现象的出现会使得大模型难以应用于面向特定任务、高度依赖确定性的领域, Ji[20] 指出可以通过两类方法缓解幻觉现象, 一是数据相关的方法, 另一是模型的训练推理相关的方法。数据相关的方法通过构建可信的数据集、对数据集进行自动化清洗、信息增强等手段抑制数据的源-引用分歧现象; 模型训练与推理相关的方法是指通过进行增强训练、后训练, 改造 Encoder-Decoder 及 Attention 架构来缓解大模型中的幻觉现象。

5 自然语言生成评估

5.1 BLEU

BLEU(bilingual evaluation understudy) 双语替换评测是用于自然语言生成的文本质量的一种方法, 其原理是将机器翻译得到的文本与专业人士翻译的文本进行比较, 其相似性越高则说明机器翻译文本质量表现好。BLEU 是自然语言生成领域一种流行的自动化评估方法。

BLEU 方法由 Papineni[21] 提出, 这种方法大大降低了评估机器翻译的重复工作。BLEU 方法首先需要计算精度:

$$p_n = \frac{\sum \sum \text{Count}_{\text{clip}}(n - \text{gram})}{\sum \sum \text{Count}(n - \text{gram})}$$

BLEU 分数最终由下式计算得到,

$$\text{BLEU} = \text{BP} * \exp\left(\sum_{n=1}^N w_n \log p_n\right)$$

其中 w_n 为权重因子, BP 为惩罚因子, 当候选译文不足参考译文时惩罚。

$$\text{BP} = \begin{cases} 1 & \text{if } c > r \\ e^{1-\frac{r}{c}} & \text{if } c \leq r \end{cases}$$

c 为候选译文长度, r 为所有参考译文中与候选译文长度最接近的长度。

6 总结

本次调研报告从自然语言生成方法、自然语言生成应用、自然语言生成评估等方面进行调研, 此外还调研了大模型出现以来, 自然语言生成中出现的幻觉现象。调研过程主要参考前人研究者们总结的综述论文[22] 以及经典论文, 自然语言生成方法方面从传统的模版生成方法到以神经网络为基础的生成方法, 调研了从最初的 RNN 到 Transformer 架构的演变。自然语言生成应用方面主要调研了自然语言生成在机器翻译、对话系统、报告生成、文章摘要等方面的应用及相关工作, 自然语言生成幻觉方面主要以 Ji[20] 工作为基础调研了大模型幻觉的定义、出现原因及缓解方法, 最后调研了自然语言生成中流行的 BLEU 自动化评估方法。

引用

- [1] E. Reiter and R. Dale, "Building applied natural language generation systems", Natural Language Engineering, vol. 3, no. 1, pp. 57 – 87, 1997.
- [2] Y. Bengio, R. Ducharme, P. Vincent, and C. Jauvin, "A neural probabilistic language model. journal of machine learning research, Vol. 3, No". Feb, 2003.
- [3] S. Mcroy, S. Channarukul, and S. Ali, "An augmented template-based approach to text realization", Natural Language Engineering, vol. 9, pp. 381 – 420, 2003, doi: 10.1017/S1351324903003188.
- [4] M. I. Jordan, "Serial order: A parallel distributed processing approach", Advances in psychology, vol. 121. Elsevier, pp. 471 – 495, 1997.
- [5] J. L. Elman, "Finding structure in time", Cognitive science, vol. 14, no. 2, pp. 179 – 211, 1990.
- [6] S. Hochreiter and J. Schmidhuber, "Long short-term memory", Neural computation, vol. 9, no. 8, pp. 1735 – 1780, 1997.
- [7] K. Cho et al., "Learning phrase representations using RNN encoder-decoder for statistical machine translation", arXiv preprint arXiv:1406.1078, 2014.
- [8] I. Sutskever, O. Vinyals, and Q. V. Le, "Sequence to sequence learning with neural networks", Advances

- in neural information processing systems, vol. 27, 2014.
- [9] D. Bahdanau, K. Cho, and Y. Bengio, “Neural machine translation by jointly learning to align and translate”, arXiv preprint arXiv:1409.0473, 2014.
 - [10] A. Vaswani et al., “Attention is all you need”, Advances in neural information processing systems, vol. 30, 2017.
 - [11] A. Radford, K. Narasimhan, T. Salimans, I. Sutskever, and others, “Improving language understanding by generative pre-training”, 2018.
 - [12] F. Stahlberg, “Neural machine translation: A review”, Journal of Artificial Intelligence Research, vol. 69, pp. 343 – 418, 2020.
 - [13] H. Schwenk, “Continuous space translation models for phrase-based statistical machine translation”, in Proceedings of COLING 2012: Posters, 2012, pp. 1071 – 1080.
 - [14] Z. Wei et al., “Task-oriented dialogue system for automatic diagnosis”, in Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers), 2018, pp. 201 – 207.
 - [15] W. Wang, Z. Zhang, J. Guo, Y. Dai, B. Chen, and W. Luo, “Task-oriented dialogue system as natural language generation”, in Proceedings of the 45th International ACM SIGIR Conference on Research and Development in Information Retrieval, 2022, pp. 2698 – 2703.
 - [16] W. Boag, T.-M. H. Hsu, M. McDermott, G. Berner, E. Alesentzer, and P. Szolovits, “Baselines for chest x-ray report generation”, in Machine learning for health workshop, 2020, pp. 126 – 140.
 - [17] S. Varges, H. Bieler, M. Stede, L. C. Faulstich, K. Irsig, and M. Atalla, “SemScribe: Natural Language Generation for Medical Reports.”, in LREC, 2012, pp. 2674 – 2681.
 - [18] A. M. Rush, S. Chopra, and J. Weston, “A neural attention model for abstractive sentence summarization”, arXiv preprint arXiv:1509.00685, 2015.
 - [19] R. Nallapati, F. Zhai, and B. Zhou, “Summarunner: A recurrent neural network based sequence model for extractive summarization of documents”, in Proceedings of the AAAI conference on artificial intelligence, 2017.
 - [20] Z. Ji et al., “Survey of hallucination in natural language generation”, ACM Computing Surveys, vol. 55, no. 12, pp. 1 – 38, 2023.
 - [21] K. Papineni, S. Roukos, T. Ward, and W.-J. Zhu, “Bleu: a method for automatic evaluation of machine translation”, in Proceedings of the 40th annual meeting of the Association for Computational Linguistics, 2002, pp. 311 – 318.
 - [22] A. Gatt and E. Krahmer, “Survey of the State of the Art in Natural Language Generation: Core tasks, applications and evaluation”. 2018.