

文本摘要简述

冯夏冲

1 简介

随着互联网产生的文本数据越来越多，文本信息过载问题日益严重，对各类文本进行一个“降维”处理显得非常必要，文本摘要便是其中一个重要的手段。文本摘要旨在将文本或文本集合转换为包含关键信息的简短摘要。

文本摘要按照输入类型可分为单文档摘要（Single document）和多文档摘要（Multiple document）。单文档摘要从给定的一个文档中生成摘要，多文档摘要从给定的一组主题相关的文档中生成摘要。与单文档摘要相比，多文档摘要面临的减小句子冗余度、确定句子顺序、确定压缩比率（从每个文档中抽取句子的比例）、指代消解问题都更加的突出。按照输出类型可分为抽取式摘要（Extractive）和生成式摘要（Abstractive）。抽取式摘要从源文档中抽取关键句和关键词组成摘要，摘要全部来源于原文。生成式摘要根据原文，允许生成新的词语、短语来组成摘要。按照目的可以分为查询无关摘要（Generic）和查询相关摘要（Query focused）。查询无关摘要是对一个文档做整体性摘要，而查询相关摘要是根据用户查询语句表达的信息需求（Information need）来对一篇文档做出摘要总结。按照有无监督数据可以分为有监督摘要（Supervised）和无监督摘要（Unsupervised）。

本文主要关注单文档、查询无关、有监督、抽取式、生成式摘要。对目前常见的文本摘要方法进行简要概述。

2 抽取式摘要

抽取式方法从原文中选取关键词、关键句组成摘要。这种方法天然的在词法、语法、句法上错误率低，保证了一定的效果。传统的抽取式摘要方法使用图方法、聚类等方式完成无监督摘要。目前流行的基于神经网络的抽取式摘要往往将问题形式化为序列标注（Sequence labeling）和句子排序（Sentence ranking）两类任务。下面首先介绍传统的抽取式摘要方法，接着分类简述基于神经网络的抽取式摘要方法。

2.1 传统方法

2.1.1 Lead-3

一般来说，好的作者常常会在标题和文章开始就表达主题，因此最简单的方法就是抽取文章中的前几句作为摘要。常用的方法为 Lead-3，即抽取文章的前三句作为文章的摘要。Lead-3 方法虽然简单直接，但却是非常有效的方法。

2.1.2 TextRank

TextRank 算法仿照 PageRank，将句子作为节点，使用句子间相似度，构造无向有权边。使用边上的权值迭代更新节点值，最后选取 N 个得分最高的节点，作为摘要。算法流程如图 1 所示。

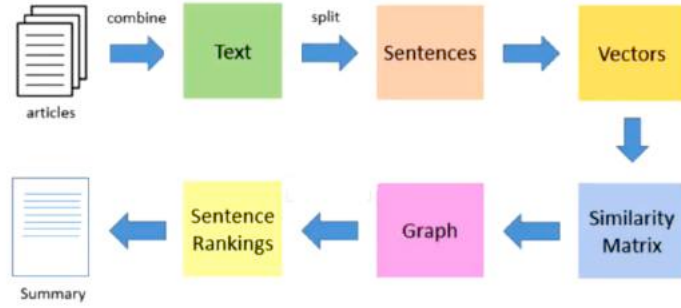


图 1: TextRank 算法流程

2.1.3 聚类

将文章中的句子视为一个点，按照聚类的方式完成摘要。例如 [Padmakumar and Saran] 等人将文章中的句子使用 Skip thought vectors 和 Paragraph embeddings 两种方式进行编码，得到句子级别的向量表示，再使用 K 均值聚类和 Mean-Shift 聚类进行句子聚类，得到 N 个类别。最后从每个类别中，选择距离质心最近的句子，得到 N 个句子，作为最终摘要。

2.2 序列标注方式

抽取式摘要可以建模为序列标注 (Sequence labeling) 任务进行处理，其核心想法是：为原文中的每一个句子打一个二分类标签 (0 或 1)，0 代表该句不属于摘要，1 代表该句属于摘要。最终摘要由所有标签为 1 的句子构成。

2.2.1 序列标注摘要基本框架

将文本摘要建模为序列标注任务的关键在于获得句子的表示，即将句子编码为一个向量，根据该向量进行二分类任务，例如 AACL17 中，[Nallapati et al., 2017] 等人的工作，使用双向 GRU 分别建模词语级别和句子级别的表示。模型 SummaRuNNer 如图 2所示。蓝色部分为词语级别表示，红色部分为句子级别表示，对于每一个句子表示，有一个 0、1 标签输出，指示其是否是摘要。

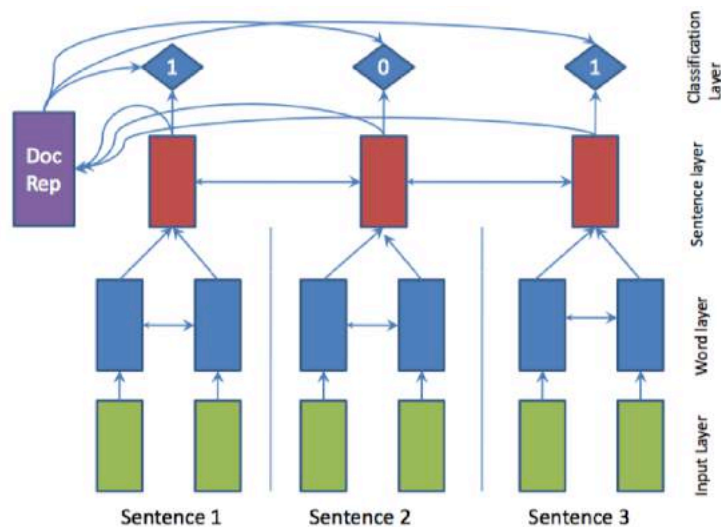


图 2: SummaRuNNer 模型

该模型的训练需要监督数据，现有数据集往往没有对应的句子级别的标签，因此需要通过启发式规则进行获取。具体方法为：首先选取原文中与标准摘要计算 ROUGE 得分最高的一句话加入候选集合，接着继续从原文中进行选择，保证选出的摘要集合 ROUGE 得分增加，直至无法满足该条件。得到的候选摘要集合对应的句子设为 1 标签，其余为 0 标签。

2.2.2 序列标注结合多任务学习

在 EMNLP17 中，[Isonuma et al., 2017] 等人结合文档分类任务来辅助完成抽取式摘要任务。模型如图 3 所示。Sentence Encoder 部分，使用 CNN 作为特征抽取器，得到句子级别的表示。采用 RNN Encoder-Decoder 框架，输入句子的表示，然后在 Decoder(Sentence Extraction) 的部分，每一步预测一个概率，代表该句是摘要句的概率。将这个概率和 Encoder 部分的句子表示相乘，得到一个文档的表示。使用该文档的表示来预测文档类别。

其核心想法在于将文档的主题作为一种粗糙的摘要，如果一个模型可以预测出文章的主题，说明其具备了从原文中抽取与主题相关句的能力，而摘要又往往由主题相关句构成。因此使用文档分类任务和句子抽取任务相互辅助，提高模型抽取关键句的能力。

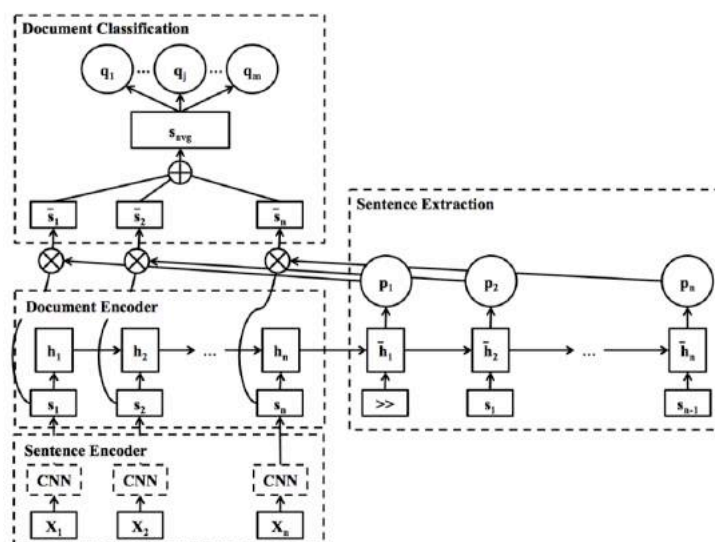


图 3: 文档分类和句子抽取多任务模型

该框架使用了 baby step 课程学习 (curriculum learning)，将数据集分为了三个部分，第一部分与第二部分数据相同，第一部分用来训练句子抽取任务，第二部分用来训练文档分类任务，第三部分没有真实的摘要，将文档的主题类别作为假设的摘要，进行训练。最终摘要仅使用 Sentence Extraction 部分，根据每一步预测的概率产生。

数据集中的文档类别可以根据新闻的类别或者自定义类别得到，句子级别的标签使用启发式规则得到。

2.2.3 序列标注结合 Seq2Seq

抽取式摘要还可以在序列标注的基础上结合 Seq2Seq 和强化学习完成。

ACL18 中，[Zhang et al., 2018] 等人在序列标注的基础上，使用 Seq2Seq 学习一个句子压缩 (Sentence compression) 模型，使用该模型来衡量选择句子的好坏，并结合强化学习完成模型训练。其模型 Latent 如图 4 所示。

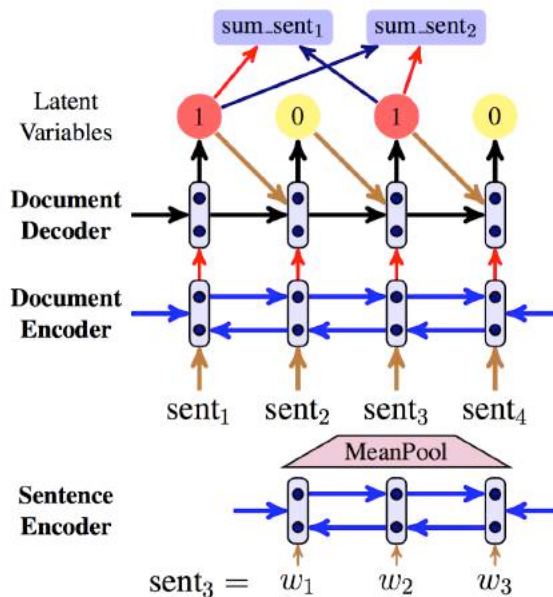


图 4: Latent 模型

该方法的核心关注点是：摘要数据集往往没有对应的句子级别的标签，需要通过启发式规则获取，然而仅仅利用这些标签训练模型会丢失很多标准摘要中重要的信息。因此 Latent 模型不采用序列标注方法计算标签级别的损失来训练模型，而是将序列标注作为中间的步骤。在得到序列标注的概率分布之后，从中采样候选摘要集合，与标准摘要对比计算损失，可以更好地利用标准摘要中的信息。

Sentence Encoder、Document Encoder、Document Decoder 与之前模型基本相同，用于获得句子级别的表示，在得到 0、1 标签的概率分布之后可以从中采样，获得候选摘要集合。利用训练好的 Seq2Seq 模型可以计算从候选摘要到真实摘要的生成概率和从真实摘要到候选摘要的生成概率，将两种概率结合作为强化学习中的奖励 Reward 训练模型。

2.3 Seq2Seq 方式

之前的抽取式摘要往往建模句子级别的表示，而忽略了关键词的作用。ACL18 中，[Jadhav and Rajan, 2018] 等人直接使用 Seq2Seq 模型来交替生成词语和句子的索引序列来完成抽取式摘要任务。其模型 SWAP-NET 如图 5 所示。

在解码的每一步，计算一个 Switch 概率指示生成词语或者句子。最后解码出的是词语和句子的混合序列。最终摘要由产生的句子集合选出。除了考虑生成句子本身的概率之外，还需要考虑该句是否包含了生成的词语，如果包含，则得分高，最终选择 top k 句作为摘要。

2.4 句子排序方式

抽取式摘要还可以建模为句子排序 (Sentence ranking) 任务完成，与序列标注任务的不同点在于，序列标注对于每一个句子表示打一个 0、1 标签，通常使用 *sigmoid* 完成，而句子排序任务则是针对每个句子输出其是否是摘要句的概率，将概率经过 *softmax* 归一化，最终依据概率，选取 top k 个句子作为最终摘要。虽然任务建模方式 (最终选取摘要方式) 不同，但是其核心关注点都是对于句子的建模。

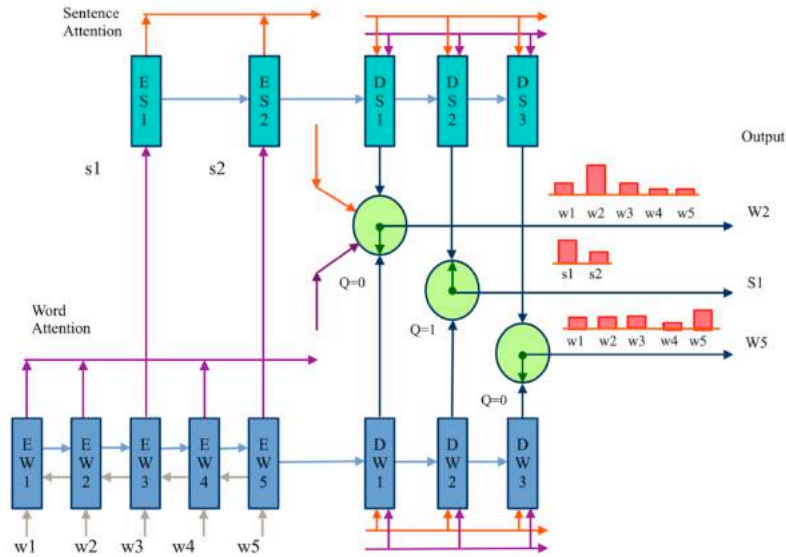


图 5: SWAP-NET 模型

2.4.1 句子排序结合强化学习

在 NAACL18 中, [Narayan et al., 2018] 等人将摘要抽取作为一种句子排序工作, 整体模型 REFRESH 如图 6所示。

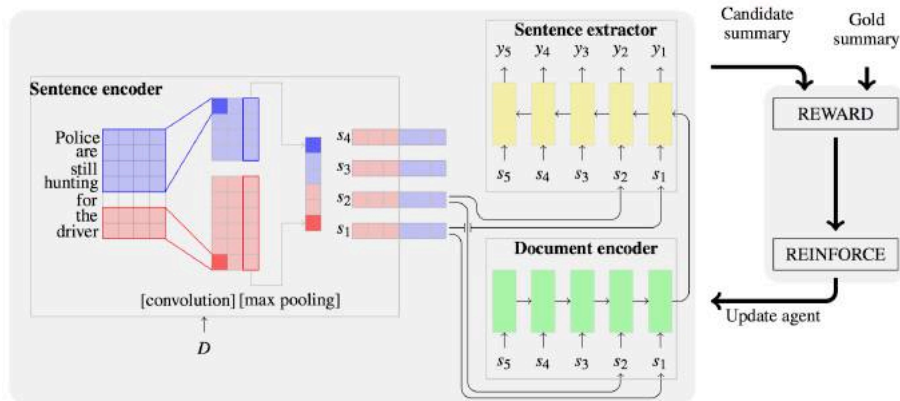


图 6: REFRESH 模型

Encoder 部分使用多核 CNN 作为句子级别的特征抽取器。之后将句子表示送入 Encoder-Decoder 给每一个句子打一个标签以及概率, 指示每个句子是否是摘要。将每个句子对应的概率经过 *softmax* 进行归一化以后, 选取 top k 个句子作为最终摘要。

这篇工作没有采用交叉熵 Cross-entropy 作为目标函数, 而是使用强化学习引入 ROUGE 作为奖励, 使用强化学习的方法进行参数更新。由于 REINFORCE 需要有采样过程, 为了缩小搜索空间, 候选摘要并不是从整个空间中搜索, 而是首先选取 p 个单个 ROUGE 高分的句子, 进行组合, 再从组合的结果中选择 k 个 ROUGE 高的摘要作为搜索空间, 计算期望奖励。

2.4.2 句子排序结合新的打分方式

之前的模型都是在得到句子的表示以后对于句子进行打分, 这就造成了打分与选择是分离的, 先打分, 后根据得分进行选择。没有利用到句子之间的关系。在 ACL18 中, [Zhou et al., 2018] 等

人提出了一种新的打分方式，使用句子受益作为打分方式，考虑到了句子之间的相互关系。其模型 NeuSUM 如图 7 所示。

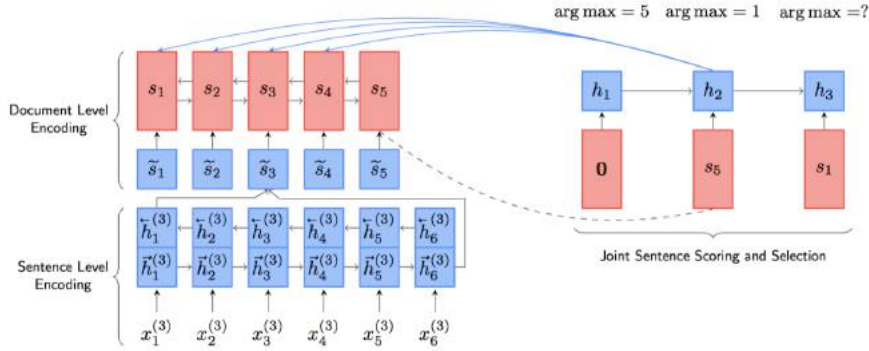


图 7: NeuSUM 模型

句子编码部分与之前基本相同。打分和抽取部分使用单向 GRU 和双层 MLP 完成。单向 GRU 用于记录过去抽取句子的情况，双层 MLP 用于打分。打分如公式 1 所示。

$$g(S_t|S_{t-1}) = r(S_{t-1} \cup \{S_t\}) - r(S_{t-1}) \quad (1)$$

其中 r 代表 ROUGE 评价指标， S_{t-1} 代表已经选择的句子集合， S_t 代表候选句子，目标是使得 g 越大越好，即选择最大化收益的句子。

因此在打分和选择部分，逐步选择使得 g 最高的句子，直到无法满足该条件或者达到停止条件为止。集合 S 为最终摘要。

3 生成式摘要

抽取式摘要在语法、句法上有一定的保证，但是也面临了一定的问题，例如：内容选择错误、连贯性差、灵活性差等问题。生成式摘要允许摘要中包含新的词语或短语，灵活性高，随着近几年神经网络模型的发展，序列到序列 (Seq2Seq) 模型被广泛的用于生成式摘要任务，并取得一定的成果。

仅使用 Seq2Seq 来完成生成式摘要存在如下问题：(1) 重复描述某些事实性信息，(2) 未登录词问题 (OOV)，(3) 生成重复。现在被广泛应用于生成式摘要的框架由 [See et al., 2017] 等人在 ACL17 中提出，在基于注意力机制 (Attention mechanism) 的 Seq2Seq 基础上增加了 Copy 和 Coverage 机制，有效的缓解了上述问题。其模型 pointer-generator 网络如图 8 所示。

其模型基本部分为基于注意力机制的 Seq2Seq 模型，使用每一步解码的隐层状态与编码器的隐层状态计算权重，最终得到 context 向量，利用 context 向量和解码器隐层状态计算输出概率。

利用 Copy 机制，需要在解码的每一步计算拷贝 (pointer) 或生成 (generator) 的概率，因为词表是固定的，该机制可以选择从原文中拷贝词语到摘要中，有效的缓解了未登录词 (OOV) 的问题。

利用 Coverage 机制，需要在解码的每一步考虑之前步的 attention 权重，结合 coverage 损失，避免继续考虑已经获得高权重的部分。该机制可以有效缓解生成重复的问题。

基于该框架可以做出一些改进，在 ICLR18 中，[Paulus et al., 2017] 等人，在该框架的基础上又使用解码器注意力机制结合强化学习来完成生成式摘要。

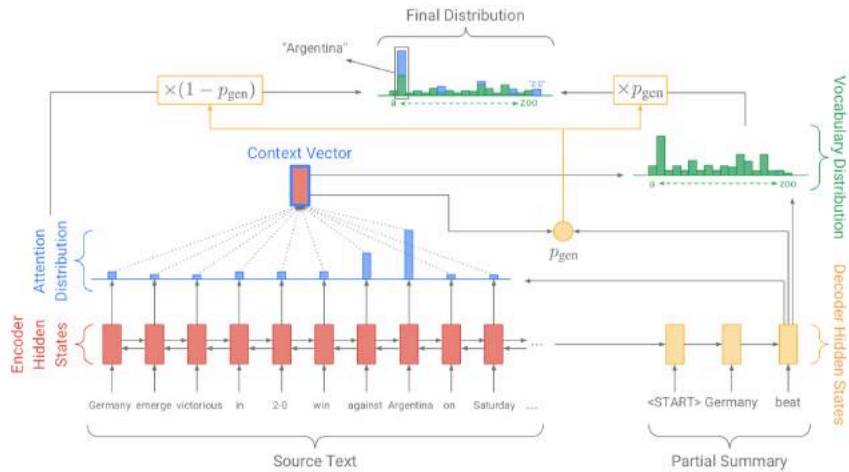


图 8: Pointer-Generator 模型

基于上述 Coverage 机制, 在 EMNLP18 中, [Li et al., 2018c] 等人基于句子级别的注意力机制, 使用句子级别的 Coverage 来使得不同的摘要句可以关注不同的原文, 缓解了生成信息重复的问题

3.1 利用外部信息

除上述问题以外, 基于 Seq2Seq 的模型往往对长文本生成不友好, 对于摘要来说, 更像是一种句子压缩 (Sentence compress), 而不是一种摘要。因此在 ACL18 中, [Cao et al., 2018] 等人, 使用真实摘要来指导文本摘要的生成。其核心想法在于: 相似句子的摘要也具有一定相似度, 将这些摘要作为软模板, 作为外部知识进行辅助。其模型 Re3 Sum 一共包含 Retrieve、Rerank、Rewrite 三个部分。

Retrieve 部分主要检索相似句子, 获得候选摘要, 其过程如图 9所示。X 是当前句子, X1-3 是训练集中的句子, 分别有 r1-3 三个摘要, 首先根据句子相似度选取 X1-3 中的句子, 然后获得摘要作为候选模板集合。

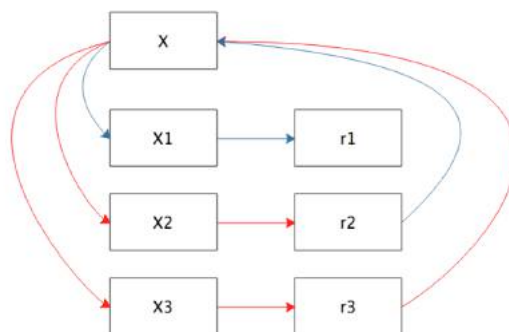


图 9: Retrieve 步骤

Rerank 部分用于排序候选模板, 在训练集中, 计算候选与真实摘要的 ROUGE 得分作为排序依据, 在开发集与测试集中, 使用神经网络计算得分作为排序依据。训练过程中, 使得预测得分尽可能与真实得分一致。

Rewrite 部分, 结合候选模板与原文生成摘要。Rerank 和 Rewrite 联合训练如图 10所示。

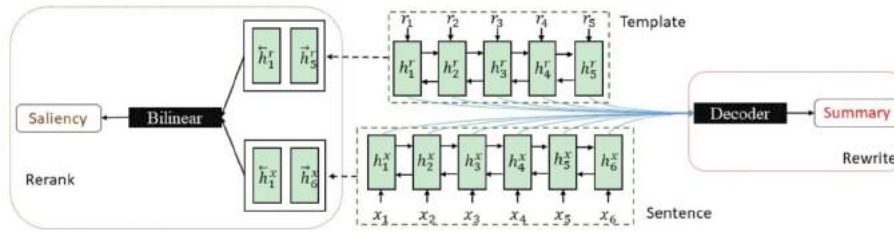


图 10: Rerank 和 Rewrite 联合训练

3.2 多任务学习

除了将本身数据集的信息作为一种外部知识以外，在 ACL18 中，[Guo et al., 2018] 等人将摘要生 (Summary generation, SG) 成作为主任务，问题生 (Question generation, QG) 成、蕴含生成 (Entailment generation, EG) 作为辅助任务进行多任务学习。问题生成任务需要根据给定的文本和答案生成问题，要求模型具有选择重要信息的能力，蕴含生成任务要求根据给定文本，有逻辑的推出输出文本，要求模型具有逻辑推理能力。在文本摘要中，定位原文中的关键信息是核心问题，根据原文生成摘要又要求模型具有一定的逻辑推理能力，使得生成的摘要与原文不违背，无矛盾。其模型 SG+QG+EG 如图 11所示。

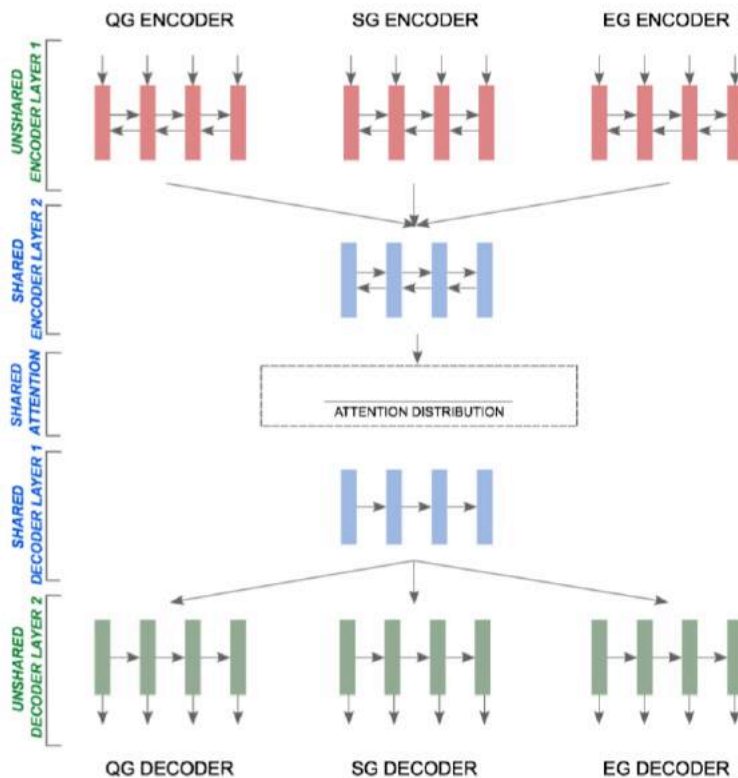


图 11: SG+TG+EG 模型

3.3 生成对抗方式

在 AAAI18 中，[Liu et al., 2018] 等人利用 SeqGAN[Yu et al., 2017] 的思想，利用生成模型 G 来生成摘要，利用判别模型 D 来区分真实摘要与生成摘要。使用强化学习的方法，更新参数。

4 抽取生成式摘要

抽取式、生成式摘要各有优点，为了结合两者的优点，一些方法也同时使用抽取结合生成的方法来完成摘要任务。

在生成式摘要中，生成过程往往缺少关键信息的控制和指导，例如 pointer-generator 网络在 copy 的过程中，无法很好地定位关键词语，因此一些方法首先提取关键内容，再进行摘要生成。

4.1 Hard 方式

从直觉上来讲，摘要任务可以大致分为两步，首先选择重要内容，其次进行内容改写。在 EMNLP18 中，[Gehrmann et al., 2018] 等人基于这种想法，提出了“Bottom Up”方式的摘要，首先使用“content selector”选择关键信息，其次使用 pointer-generator 网络生成摘要。其整体流程如图 12 所示。

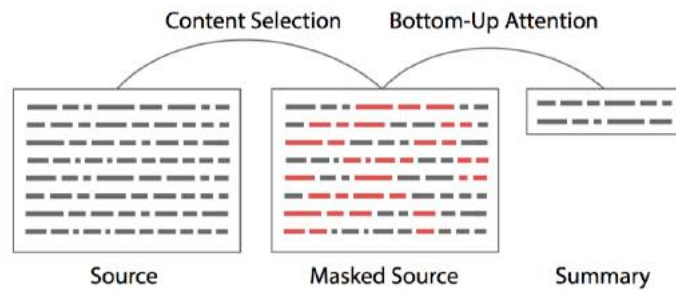


图 12: Bottom-Up 摘要流程

内容选择部分建模为词语级别序列标注任务，该部分的训练数据通过将摘要对齐到文档，得到词语级别的标签。摘要生成部分使用 pointer-generator 网络，使用内容选择部分计算的概率该修改原本 attention 概率，使得 Decoder 仅关注选择的内容。

除了上述以序列标注方式来选择关键词的方法以外，在 NAACL18 中，[Li et al., 2018a] 等人使用 TextRank 算法获得关键词，之后使用神经网络获得关键词语的表达，并将该表示结合 pointer-generator 网络生成摘要。其模型 KIGN 如图 13 所示。

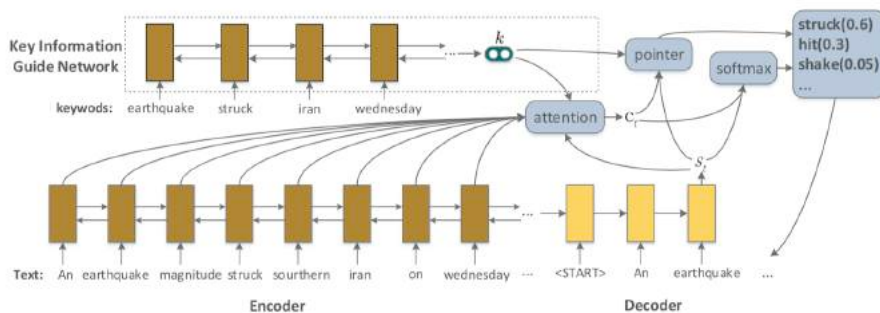


图 13: KIGN 模型

4.2 Soft 方式

上述方法从原文中选择重要的部分用来指导摘要的生成，显示的利用了文本级别的信息，属于一种 Hard 方式，在 EMNLP18 中，[Li et al., 2018b] 等人，使用门控机制，从编码得到的向量表示中选择有用的信息用于之后的摘要生成，属于一种 Soft 方式。在使用层次化 encoder 得到句子级别的向量表示之后，使用门控机制，得到新的句子级别向量，表示从中选择有用信息。其模型 InfoSelection 如图 14 所示。

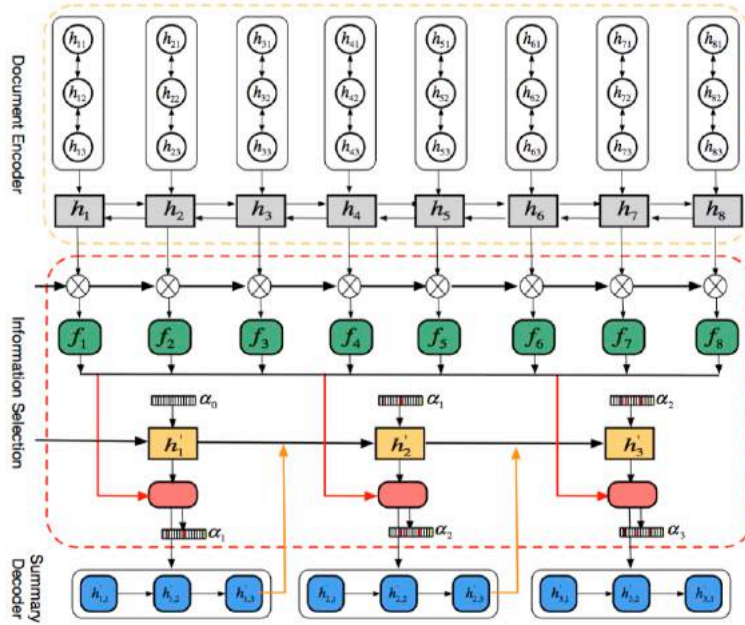


图 14: InfoSelection 模型

4.3 Soft+Hard 方式

在 ACL18 中，[Hsu et al., 2018] 等人将抽取式模型的输出概率作为句子级别的 attention 权重，用该权重来调整生成式模型中的词语级别的 attention 权重，如图 15 所示，其核心想法为：当词语级别的 attention 权重高时，句子级别的 attention 权重也高。基于此想法提出了 Inconsistency 损失函数，使得模型输出的句子级别的权重和词语级别的权重尽量一致。在最终训练时，首先分别预训练抽取式和生成式模型，之后有两种方式来结合两个模型，Hard 方式：将抽取式模型抽出的关键句直接作为生成式模型的输入；Soft 方式：将抽取式模型的输出概率用来调整词语级别的权重。

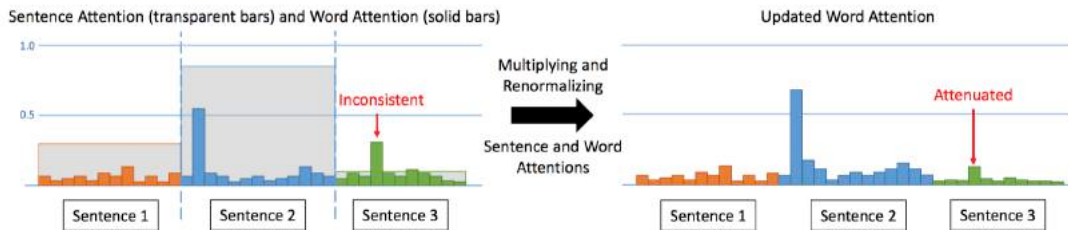


图 15: 权重调整过程

5 AAI19

在 AAI19 中, [Gao et al., 2018] 等人提出了“关注读者的生成式文本摘要 (Reader-aware abstractive text summarization)”任务, 并提出了新的数据集, 数据集示例如图 16 所示。由文档、摘要、评论构成, 共 863826 训练样例, 文档平均 67.08 个词, 摘要平均 16.56 个词, 评论平均 16.61 个词, 每个文档平均 9.11 条评论。

document	徐麟表示, 中央网信办将提供政策支持建立健全国有资本进入培育互联网企业, 完善互联网企业国内上市等相关政策; 通过在新闻网站核发新闻记者证, 开展从业人员教育培训等措施, 努力打造新媒体平台的国家队和主力军。详见长微博(Lin Xu said that the Central Network Office will provide policy support to establish a nationwide capital to enter the cultivation of internet companies, improve the policies of domestic listing of internet companies; through the press on the news website to issue a press card, carry out education and training for practitioners, and strive to create new the national team and the main force of the media platform. See Long Weibo for details.)
comments	国有资本必须介入(State-owned capital must be involved.)
	中央网信办不要把举措落实在文件上, 要切实的执行! (The Central Network Office should not implement the measures on the documents and must implement them in a practical way !)
reference	我觉得网信办要好好治治这些害群之马了(I feel that the Central Network Office should cure these black sheep.)
reference	网信办副主任: 建立健全国有资本进入培育互联网企业(Deputy director of the Central Network Office: Establishing state-owned capital to cultivate internet enterprises)

图 16: Reader-aware 数据集示例

在 AAI19 中, [Shi et al., 2018] 等人改变了之前将文档映射到摘要的做法, 而是将文档-摘要对映射到得分, 采用噪声信道模型的思想, 考虑摘要对于文档的重要性。提出了一种对比学习 (Contrastive Learning) 的方式来训练模型, 其模型 DeepChannel 如图 17 所示。

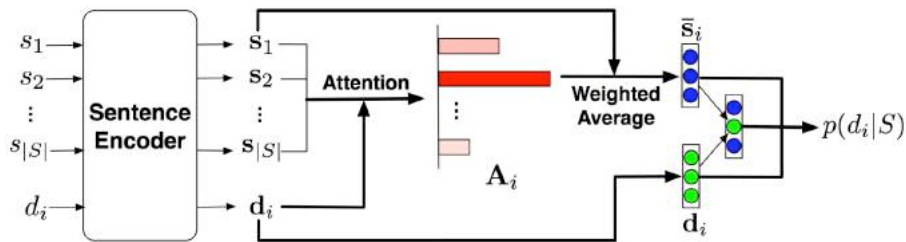


图 17: DeepChannel 结构图

得到训练好的模型之后, 贪婪地每次从原文中选择一个句子, 使得得分最高, 直达到限制条件。

6 数据集

6.1 DUC Corpus

DUC 数据集为单句摘要数据集,将长文本生成一句摘要,分为 DUC2003 和 DUC2004。DUC2003 包含 624 文档-摘要对。DUC2004 包含 500 文档-摘要对, DUC2004 文档平均 35.6 个词, 摘要平均 10.4 个词。

由于 DUC 数据集规模较小, 因此神经网络模型往往在其他数据集上进行训练, 再在 DUC 数据集上进行测试。

6.2 NYT Corpus

NYT 数据集包含了发表于 1996 至 2007 年期间的文章, 摘要由专家编写而成。该数据集的摘要有时候不是完整的句子, 并且长度较短, 大约平均 40 个词语。常用于训练“抽取式”摘要系统。

6.3 CNN/Daily Mail Corpus

目前广泛使用的是 CNN/Daily Mail 数据集, 属于多句摘要数据集, 常用于训练“生成式”摘要系统。该数据集从问答任务的数据集修改得到, 训练集包括 286817 对, 开发集包括 13368 对, 测试集包括 11487 对。其中训练数据集文章平均 766 个词, 29.74 句话, 摘要平均 53 个词, 3.72 句话。

该数据集一共有两个版本, 匿名 (Anonymized) 版本和未匿名 (Non-anonymized) 版本, 未匿名版本包括了真实的实体名 (Entity names), 匿名版本将实体使用特定的索引进行替换。

6.4 Gigaword Corpus

Gigaword 数据集属于单句摘要数据集, 摘要由文章第一句话和题目结合启发式规则构成。训练集共 3.8M, 开发集共 189K, 测试集共 1951。详细统计如图 18 所示。

Dataset	Train	Dev.	Test
Count	3.8M	189k	1951
AvgSourceLen	31.4	31.7	29.7
AvgTargetLen	8.3	8.3	8.8

图 18: Gigaword 数据统计

6.5 LCSTS Corpus

LCSTS 为中文短文本摘要数据集, 由新浪微博构建得到。如图 19 所示。

7 评价指标

Rouge(Recall-Oriented Understudy for Gisting Evaluation), 一种基于召回率的相似性度量方法, 是评估自动文摘以及机器翻译的一组指标, 考察翻译的充分性和忠实性。它通过将自动生成的摘要或翻译与一组参考摘要 (通常是人工生成的) 进行比较计算, 得出相应的分值, 以衡量自动生成的

【江西高考被曝替考 有关考生已被警方控制】人民日报记者吴齐强消息，江西高考被曝光替考，7日中午江西省教育厅发布消息称，接到有人组织替考的举报后，江西省教育厅、江西省教育考试院立即部署南昌市教育考试院，联合南昌市警方开展调查核实，有关考生已被警方控制。调查进展情况将及时向社会公布。

图 19: LCSTS 构建示例

摘要或翻译与参考摘要之间的“相似度”。用于文本摘要的评价指标主要有 ROUGE-1、ROUGE-2、ROUGE-L 三个指标，其计算分别涉及 Uni-gram、Bi-gram 和 Longest common sub-sequence。

8 模型总结

表格 1 简要概括了上述模型的核心特点。

9 实验结果

表格 2 总结了上述模型在 CNN/Daily Mail 数据集上的实验结果。

10 总结

文本摘要作为传统的自然语言处理任务，至今依旧有新的发展和 innovation，这一方面得益于模型、方法、语料的支撑，另一方面也是由于摘要任务自身的重要性。摘要生成作为文本生成的一种，除了有着重复、冗余、不连贯、生成较短等问题，还有着摘要任务特定的问题，其中最核心的为：如何确定关键信息。当下的文本摘要更关注“什么是真正的摘要”，而不仅仅是简单地句子压缩。利用外部知识，利用关键词信息等方式来更好的辅助摘要的生成。各个摘要模型各有优点，在评价指标上各有优势。因此在考虑评价指标的同时，更应该关注问题本身，才能使得方法更具有普适性，更好的满足真实需要。

参考文献

- Ziqiang Cao, Wenjie Li, Sujian Li, and Furu Wei. Retrieve, rerank and rewrite: Soft template based neural summarization. In *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, volume 1, pages 152–161, 2018.
- Shen Gao, Xiuying Chen, Piji Li, Zhaochun Ren, Lidong Bing, Dongyan Zhao, and Rui Yan. Abstractive text summarization by incorporating reader comments. *arXiv preprint arXiv:1812.05407*, 2018.
- Sebastian Gehrmann, Yuntian Deng, and Alexander M Rush. Bottom-Up Abstractive Summarization. *EMNLP*, 2018.
- Han Guo, Ramakanth Pasunuru, and Mohit Bansal. Soft layer-specific multi-task summarization with entailment and question generation. *arXiv preprint arXiv:1805.11004*, 2018.

模型名	会议	类型	核心想法
SummaRuNNer	AAAI17	抽取式	神经网络句子表示建模
MultiTask	EMNLP17	抽取式	文档分类、句子抽取
Latent	ACL18	抽取式	使用 Seq2Seq 计算损失，充分利用标准摘要信息，而非句子级别的标签损失
SWAP-NET	ACL18	抽取式	利用 Switch 机制，生成词语和句子的索引序列，根据生成词语来选择最终摘要
REFRESH	NAACL18	抽取式	强化学习
NeuSUM	ACL18	抽取式	句子受益作为新的打分方式
DeepChannel	AAAI19	抽取式	将文档-摘要对映射到得分
Pointer-generaor	ACL17	生成式	注意力 Seq2Seq、Copy 机制、Coverage 机制
Re ³ Sum	ACL18	生成式	利用相似句子摘要作为外部知识和软模板
GAN	AAAI18	生成式	利用 SeqGAN 的思想完成摘要任务
Bottom-up	EMNLP18	抽取 + 生成	使用内容选择器选择关键信息，并修改注意力机制
KIGN	NAACL18	抽取 + 生成	使用 TextRank 算法首先获得关键词
InfoSelection	EMNLP18	抽取 + 生成	使用门控机制从表示中提取关键信息
Inconsistency loss	ACL18	抽取 + 生成	使用抽取式模型输出概率来调整词语级别的 attn 权重

表 1: 模型总结

模型名	会议	类型	R-1	R-2	R-L	数据集类型
Lead-3		抽取	39.2	15.7	35.5	匿名
Lead-3		抽取	40.34	17.70	36.57	未匿名
SummaRuNNer	AAAI17	抽取	39.6	16.2	35.3	匿名
Latent	ACL18	抽取	41.05	18.77	37.54	未匿名
SWAP-NET	ACL18	抽取	41.6	18.3	37.7	匿名
REFRESH	NAACL18	抽取	40.0	18.2	36.6	未匿名
NeuSUM	ACL18	抽取	41.59	19.01	37.98	未匿名
DeepChannel	AAAI19	抽取	41.50	17.77	37.62	未匿名
Pointer-generaor	ACL17	生成	39.53	17.28	36.38	未匿名
SG+EG+QG	ACL18	生成	39.81	17.64	36.54	未匿名
GAN	AAAI18	生成	39.92	17.65	36.71	匿名
Bottom-up	EMNLP18	抽取 + 生成	41.22	18.68	38.34	未匿名
KIGN	NAACL18	抽取 + 生成	38.95	17.12	35.68	匿名
InfoSelection	EMNLP18	抽取 + 生成	41.54	18.18	36.47	未匿名
Inconsistency loss	ACL18	抽取 + 生成	40.68	17.97	37.13	未匿名

表 2: 实验结果总结

- Wan-Ting Hsu, Chieh-Kai Lin, Ming-Ying Lee, Kerui Min, Jing Tang, and Min Sun. A unified model for extractive and abstractive summarization using inconsistency loss. *arXiv preprint arXiv:1805.06266*, 2018.
- Masaru Isonuma, Toru Fujino, Junichiro Mori, Yutaka Matsuo, and Ichiro Sakata. Extractive summarization using multi-task learning with document classification. In *Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing*, pages 2101–2110, 2017.
- Aishwarya Jadhav and Vaibhav Rajan. Extractive summarization with swap-net: Sentences and words from alternating pointer networks. In *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, volume 1, pages 142–151, 2018.
- Chenliang Li, Weiran Xu, Si Li, and Sheng Gao. Guiding generation for abstractive text summarization based on key information guide network. In *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 2 (Short Papers)*, volume 2, pages 55–60, 2018a.
- Wei Li, Xinyan Xiao, Yajuan Lyu, and Yuanzhuo Wang. Improving neural abstractive document summarization with explicit information selection modeling. In *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing*, pages 1787–1796, 2018b.
- Wei Li, Xinyan Xiao, Yajuan Lyu, and Yuanzhuo Wang. Improving neural abstractive document summarization with structural regularization. In *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing*, pages 4078–4087, 2018c.
- Linqing Liu, Yao Lu, Min Yang, Qiang Qu, Jia Zhu, and Hongyan Li. Generative adversarial network for abstractive text summarization. In *Thirty-Second AAAI Conference on Artificial Intelligence*, 2018.
- Ramesh Nallapati, Feifei Zhai, and Bowen Zhou. Summarunner: A recurrent neural network based sequence model for extractive summarization of documents. In *Thirty-First AAAI Conference on Artificial Intelligence*, 2017.
- Shashi Narayan, Shay B Cohen, and Mirella Lapata. Ranking sentences for extractive summarization with reinforcement learning. *arXiv preprint arXiv:1802.08636*, 2018.
- Aishwarya Padmakumar and Akanksha Saran. Unsupervised text summarization using sentence embeddings.
- Romain Paulus, Caiming Xiong, and Richard Socher. A Deep Reinforced Model for Abstractive Summarization. *CoRR*, 2017.
- Abigail See, Peter J Liu, and Christopher D Manning. Get to the point: Summarization with pointer-generator networks. *arXiv preprint arXiv:1704.04368*, 2017.
- Jiaxin Shi, Chen Liang, Lei Hou, Juanzi Li, Zhiyuan Liu, and Hanwang Zhang. Deepchannel: Saliency estimation by contrastive learning for extractive document summarization. *arXiv preprint arXiv:1811.02394*, 2018.

Lantao Yu, Weinan Zhang, Jun Wang, and Yong Yu. Seqgan: Sequence generative adversarial nets with policy gradient. In *Thirty-First AAAI Conference on Artificial Intelligence*, 2017.

Xingxing Zhang, Mirella Lapata, Furu Wei, and Ming Zhou. Neural latent extractive document summarization. *arXiv preprint arXiv:1808.07187*, 2018.

Qingyu Zhou, Nan Yang, Furu Wei, Shaohan Huang, Ming Zhou, and Tiejun Zhao. Neural document summarization by jointly learning to score and select sentences. *arXiv preprint arXiv:1807.02305*, 2018.