arXiv:2302.00455v1 [cs.CL] 1 Feb 2023

HunSum-1：匈牙利语抽象式摘要数据集

Botond Barta[1,2], Dorina Lakatos[1,2], AttilaNagy[1], Milán Konor Nyist[1], Judit Ács[2]{botondbarta, dorinalakatos, acsjudit}@sztaki.hu, attila.nagy234@gmail.com, nyist.milan78@gmail.com

[1]自动化与应用信息学系 布达佩斯科技与经济大学[2]厄特沃什·罗兰研究网络计算机科学与控制研究所

摘要。我们介绍 *HunSum-1*：一个匈牙利语抽象式摘要数据集，包含 114 万篇新闻文章。该数据集通过 CommonCrawl 收集、清理并去重来自 9 个匈牙利主要新闻网站的数据构建而成。我们利用该数据集，构建了基于 huBERT 和 mT5 的抽象式摘要模型。通过对模型结果进行定量和定性分析，我们展示了所创建数据集的价值。*HunSum-1* 数据集、我们实验中使用的所有模型以及代码[1] 均以开源形式提供。

1 引言

文档自动摘要（Nenkova 等，2011）是自然语言处理领域中一项受到广泛研究的任务。摘要旨在将原始文档压缩为简短精炼的文本，同时涵盖输入中的所有关键信息。一般而言，摘要有两种方法：抽取式和抽象式。抽取式方法通过明确提取相关词元或句子，直接利用输入文档生成摘要。相比之下，抽象式摘要则根据输入生成新文本。从计算角度来看，抽取式摘要方法通常比抽象式方法复杂度更低，但也存在一些局限，例如难以处理长句，或难以确保提取出的句子之间连贯。要生成更复杂的摘要，则需要采用抽象式摘要；在此过程中，模型可能会生成原始文档中未出现的词语。

本文介绍了*HunSum-1*，这是一个用于匈牙利语抽象式摘要的数据集，包含 114 万篇带有导语的新闻文章。第 3 节详细讨论了我们如何从 CommonCrawl 提取数据并执行一系列预处理步骤，从而得到当前版本的 *HunSum-1*。我们还在该数据集上训练了三个基线模型，并从定量和

[1] https://github.com/dorinapetra/summarization

定性两个方面评估结果。据我们所知，这是首个公开发布的、规模如此庞大的匈牙利语抽象式摘要数据集。我们希望它能为未来的匈牙利语抽象式摘要研究奠定良好基础。

2 相关工作

Nallapati 等人（2017）提出了一种基于循环神经网络的抽取式摘要模型，名为 *SummaRuNNer*。该模型的表现优于多个模型，例如基于丰富特征的逻辑分类器（Cheng 和 Lapata，2016）、基于整数线性规划的方法（Woodsend 和 Lapata，2010）以及各种基于图的模型（Parveen 等人，2015；Wan，2010）。传统的抽取式摘要模型主要包含两个步骤：句子评分和句子选择。Zhou 等人（2018）将这两个步骤结合起来，直接根据前面的句子预测当前句子的重要性。Liu（2019）提出了 Bertsum，这是一种端到端的 BERT 模型（Devlin 等人，2019），包含句间 Transformer 层。Bertsum 在 *CNN/Daily Mail*（Hermann 等人，2015）数据集上的 ROUGE-L 得分高于此前最先进的模型 Neusum。

针对抽象式摘要，Paulus 等人（2018）提出了一种基于编码器-解码器的神经内部注意力模型，以及一种结合监督学习和强化学习的新训练流程。Gehrmann 等人（2018）提出了自下而上 注意力，这是一种用于内容选择的新技术：它利用数据高效的内容选择器，从输入中确定重要短语。基于 BERT 的 PreSumm 模型（Liu 和 Lapata，2019）采用了一种新的微调方案，对编码器和解码器使用不同的优化器。当前最先进的模型 PEGASUS 由 Zhang 等人（2020）提出。在 PEGASUS 的训练过程中，输入文档中的重要句子会被屏蔽，模型必须将这些句子一并生成，作为单个输出。这一训练目标免去了对有监督摘要数据集的需求。

Raffel 等人（2020）提出了 T5（文本到文本迁移 Transformer），这是一种基于 Transformer 的编码器-解码器模型，曾在包括摘要任务在内的多个下游任务上进行预训练，但仅适用于英语。mT5（Xue 等人，2021）是 T5 模型的多语言版本。Hasan 等人（2021）对 mT5 模型进行微调，以用于多语言摘要，其结果与单语模型相比具有竞争力。

Hermann 等人（2015）将 CNN 和 Daily Mail 文章与其摘要配对，构建了一个数据集。*CNN/Dailymail* 摘要语料库包含超过 100 万个样本。Hasan 等人（2021）介绍了一个类似的数据集，名为*XL-Sum*，其中包含 44 种语言的 100 万个带注释的数据样本。他们从 BBC 新闻中提取导语作为摘要，并将文章作为输入文档。

Yang 等人（2021）从 Index 和 HVG 的文章中构建了摘要语料库，并训练了多个 BERT 模型用于抽象式摘要：多语言区分大小写的 BERT、huBERT（Nemeskey，2020）以及 HILBERT 大型模——

型（Feldmann 等人，2021）。遗憾的是，该语料库并未公开。Makrai 等人（2022）发布了一个以 huBERT 初始化的编码器-解码器模型。他们使用包含 30 万个样本的 ELTE.DH 语料库进行训练。Agócs 和 Yang（2022）使用 PreSumm 方法训练多语言和匈牙利语 BERT 模型。Yang（2022）对 BART 模型（Lewis 等人，2020）进行了微调，用于抽象式摘要。

3 摘要数据集

我们使用 Common Crawl 语料库[2]构建了一个匈牙利新闻文章数据集。该数据集可用作抽象式摘要和标题生成任务的训练数据。

3.1 Common Crawl

Common Crawl（CC）基金会会抓取网页，并定期提供可供公众免费使用的快照。该语料库包含自 2008 年以来收集的数 PB 数据。数据以 WARC 格式存储在 Amazon S3 上，可通过 HTTP 或 S3 下载。我们使用了 CommonCrawl 下载器软件包（Indig，2018年），该软件包帮助我们根据给定域名下载文章并去重。我们选取了 9 个内容丰富的大型匈牙利新闻网站，并下载了截至 2022-27 的所有 Common Crawl 数据分段中的网页。我们的筛选范围仅限于使用专用导语字段的网站。这些门户网站涵盖政治、经济、体育等多个主题。我们收集了约 94 GB 的原始 HTML 数据。

3.2 数据集

我们从每篇文章中提取了标题、导语、文章正文、创建日期以及标签（如有）。解析时，如果没有找到导语，我们就跳过该文章。我们删除了每篇文章中的所有链接、嵌入的社交媒体内容、推荐内容和图片说明，因为这些内容对模型执行摘要任务没有帮助。属于实时博客的文章会被剔除，因为它们篇幅太短，而且通常无法脱离博客的其他内容作为独立文章。我们也剔除了图片集，因为没有图片时，它们通常不包含有意义的内容。

网页抓取后，我们获得了约188万篇文章。不过，我们对数据进行了筛选，剔除了不适合作为摘要生成模型输入的文章（例如过短的文章）。表1列出了对导语和文章长度所设的限制。我们使用 quntoken 软件包[3]进行词语和句子的分词。我们还删除了导语长于文章正文的文章，共约1.7万篇。对文章字符数设置的最小值和最大值限制则剔除了2.4万篇文章。

[2] https://commoncrawl.org/ [3] https://github.com/nytud/quntoken

筛选导语句子数上限和导语词元数下限后，分别删除了1万篇和2.5万篇文章。剔除数量最多的是筛选文章正文的最小句子数，这一步删除了约30万篇文章。我们还删除了3,000篇非匈牙利语文章，这些文章是通过语言检测模型 FastText 识别出来的（Joulin 等人，2016b,a）。

约束条件	值
文章最少字符数	200
文章最多字符数	15000
导语最少词元数	6
导语最多句子数	5
文章句子最小数量	6

表1：文章和导语的长度限制。

筛选后，我们使用局部敏感哈希（LSH）对文章进行去重，以移除冗余文章。如果两篇文章的文本有 45% 相似，则视为等同。若发现两篇文章相似，则删除没有导语的一篇；如果两篇都有导语，则保留爬取时间较晚的一篇。采用这一阈值后，约有 36.3 万篇文章被移除。去重后的数据集包含 114 万个数据样本。表2展示了各新闻网站的数据样本数量。

Site	数量
24.hu	359.4k
origo.hu	305.0k
hvg.hu	216.8k
index.hu	154.5k
nepszava.hu	56.7k
portfolio.hu	22.6k
m4sport.hu	17.9k
metropol.hu	11.1k
telex.hu	4.4k

表2：每个网站的数据样本数量。

该数据集包含1999年至2022年各年份的文章。图1展示了文章在各年份的分布。尽管使用了截至2022年7月的所有 Common Crawl 数据分段，数据集中来自2022年的样本仍显著较少。这可能是因为 CC 存在一定延迟。

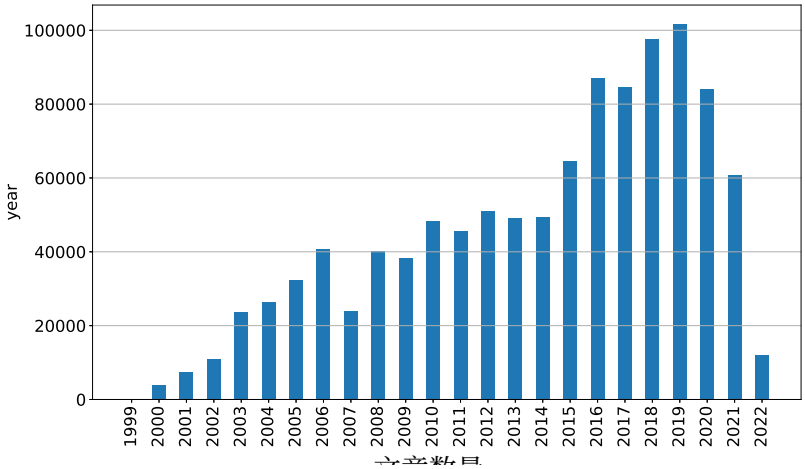

图1：按年份统计的文章数量。

表3介绍了文章和导语中的词元及句子的平均数量。*Népszava* 和 *Telex* 的文章显著长于其他网站。

	文章		Lead	
Site	**词元**	**句子**	**词元**	**句子**
24.hu	323	17	21	1.3
origo.hu	396	19	40	1.9
hvg.hu	378	17	30	1.5
index.hu	517	26	43	2.2
nepszava.hu	**869**	**39**	32	1.5
portfolio.hu	457	22	**52**	2.1
m4sport.hu	385	23	28	1.3
metropol.hu	305	17	25	1.4
telex.hu	839	38	49	**2.3**

表3：文章和导语的平均长度。

我们计算了多种自动定量指标，以评估*HunSum-1*的质量。新颖 n 元语法比例（NNG-n）（Narayan 等人，2018年）衡量摘要中未出现在原始文章中的 n 元语法所占的比例。Bommasani 和 Cardie（2020年）提出了一种压缩指标

（CMP），其定义如下：

$$\boldsymbol{CMP}(A,S) = 1 - \frac{|S|}{|A|} \tag{1}$$

其中，$|S|$ 和 $|A|$ 分别表示摘要和文章的长度。最后，我们基于 Hasan 等人（2021）计算冗余度指标（RED-n）。它定义为摘要 S 中冗余 n 元语法的数量与 n 元语法总数之比：

$$\begin{aligned} \boldsymbol{RED}(S) &= \frac{\sum_{i=1}^{m}(f_i - 1)}{\sum_{i=1}^{m} f_i} \\ &= 1 - \frac{m}{|S| - n + 1} \end{aligned} \tag{2}$$

其中，$\{f_1, f_2, ..., f_m\}$ 表示对应于唯一 n 元语法$\{g_1, g_2, ..., g_m\}$的频次。如果摘要的新颖 n 元语法比例和压缩率较高，且冗余度指标较低，通常就被认为质量较高。在*HunSum-1* 数据集上计算的上述指标见表 4。

NNG-1	NNG-2	NNG-3	压缩率	RED-1	RED-2
78.0	90.09	95.48	88.58	6.58	0.1

表 4：*HunSum-1*的内在评估。

我们创建了训练集、开发集和测试集划分，并将在 HuggingFace 数据集中心以 SZTAKI-HLT/HunSum-1 之名提供。开发集和测试集均包含 2000 篇文档，并通过从每个新闻页面进行分层抽样创建。

4 模型

我们对两种语言模型进行了实验：一种是编码器-解码器模型，其两侧均由 BERT 模型构成（Rothe 等人，2020年）。我们使用匈牙利语版本的 BERT（Devlin 等人，2019年）huBERT（Nemeskey，2020年）（Bert2Bert）的权重初始化两侧，另一种是经过预训练的多语言 mT5（Xue 等人，2021年）。

4.1 mT5

mT5 是文本到文本迁移 Transformer（T5）（Raffel 等人，2020年）的多语言版本，支持包括匈牙利语在内的 101 种语言。mT5 在 mC4 数据集上训练，该数据集包含从 Common Crawl 数据集中提取的总计 6.3 万亿词元。匈牙利语是该语料库中第 22 常见的语言，包含 3700 万个匈牙利语页面。与 T5 模型不同，mT5

并未针对任何下游任务（例如抽象式摘要）进行训练。mT5 有 5 个变体：mt5-small、mt5-base、mt5-large、mt5-xl 和 mt5-xxl，参数量分别为 3 亿、5.8 亿、12 亿、37 亿和 130 亿。由于计算资源限制，我们决定只对小型和基础变体进行实验。与 BERT 系列不同，T5 和 mT5 都是生成模型。

5 实验

我们训练了 3 个摘要模型：Bert2Bert、mT5-small 和 mT5-base。我们发现，Bert2Bert 和 mT5 模型适用的生成超参数各不相同。我们还发现，mT5 模型往往会复制输入句子，因此需要额外设置一个参数。为防止这种情况，我们将 encoder_no_repeat_ngram_size 参数设为 4。表 5 描述了训练和生成超参数。

参数	**Bert2Bert**	**mT5-小型**	**mT5-base**
批量大小	13	16	12
学习率	5e-5	5e-5	5e-5
权重衰减	0.01	0.01	0.01
预热步数	16000	3000	3000
训练轮数	15	10	10
不_重复_元语法_大小	3	3	3
num_束搜索数	5	5	5
提前_停止	True	否	否
编码器_否_重复_n 元语法_大小	-	4	4

表 5：训练和生成的超参数。

文章的最大长度为 512 个子词，生成的导语最大长度为 128 个子词。模型均在单块 NVIDIA A100 GPU 上训练。Bert2Bert 训练了 265 小时，mT5-small 训练了 54 小时，而 mT5-base 模型则训练了 119 小时。我们的模型可在 Huggingface 模型中心获取，名称分别为 SZTAKI-HLT/Bert2Bert-HunSum-1、SZTAKI-HLT/mT5-small-HunSum-1 和 SZTAKI-HLT/mT5-base-HunSum-1。

6 结果

我们使用 ROUGE 分数（面向摘要评估的召回率导向型替代评估指标）（Lin，2004）进行自动评估。ROUGE 以参考摘要为基准，计算模型输出与之之间的词语 n-gram 重叠。在我们的案例中，参考摘要就是真实文章的导语。我们计算了两个版本：一个普通版本和一个词干化版本。由于匈牙利语是一种黏着语，微小的词形变化

就可能降低 ROUGE 分数，即使摘要本身是正确的。我们使用 HuSpaCy（ Orosz 等，2022年）模型进行词干提取。表6展示了测试集上的 ROUGE 分数。

模型	原文			词干化		
	R-1	**R-2**	**R-L**	**R-1**	**R-2**	**R-L**
Bert2Bert	28.52	10.35	20.07	31.75	11.80	22.31
mT5-小型	36.49	9.50	23.48	40.86	11.71	26.25
mT5-基础	**37.70**	**11.22**	**24.37**	**41.97**	**13.65**	**27.23**
foszt2oszt[4]	23.62	5.05	16.57	26.22	5.95	18.37
hi-bart-基础[5]	29.03	8.79	19.60	32.57	10.45	21.98
hi-bart[6]	30.03	8.96	20.50	33.23	10.45	22.71
nol-bart[7]	26.38	6.92	18.05	29.25	8.06	20.09

表6：测试集上的 ROUGE 召回率分数。ROUGE-1、ROUGE-2 和 ROUGE-L 分数分别缩写为 R-1、R-2 和 R-L。

根据 ROUGE 分数，mT5-base 模型显然是最佳模型，mT5-small 则位居第二，尽管并非在所有指标上都是如此。尽管名称中有“小型”，mT5-small 并不是一个很小的模型，它有 3 亿个可训练参数。相比之下，Bert2Bert 是两个 BERT 模型的组合，只有 2.5 亿个参数。除了参数量更高之外，T5 模型是生成式模型，而 BERT 不是，这或许可以解释为什么多语言模型在这项生成任务上的表现优于单语 huBERT。

我们将自己的模型与现有的匈牙利语摘要模型进行了比较。结果见表6。根据我们测试集上的分数，我们的 mT5-base 模型优于其他匈牙利语模型。

6.1 人工评估

由于 ROUGE 分数并不总能反映生成摘要的质量，我们决定对模型进行人工评估。我们采用分层抽样，从测试集中抽取了 80 篇文章进行人工评估。我们使用 Hasan 等人（ 2021 ）的评估框架（**属性 A** 至 **C** ），评估者需要针对以下问题回答“是”或“否”。我们还新增了一道关于语法正确性的问题（**属性 D**）。

– **属性 A：** 摘要是否传达了文章的内容?

[4] https://huggingface.co/BME-TMIT/foszt2oszt [5] https://huggingface.co/NYTK/summarization-hi-bart-base-1024-hungarian [6] https://huggingface.co/NYTK/summarization-hi-bart-hungarian [7] https://huggingface.co/NYTK/summarization-nol-bart-hungarian

– **属性 B：**如果属性 A 的答案是肯定的，那么摘要是否仅包含与文章一致的信息？– **属性 C：** 如果属性 A 的答案是肯定的，那么摘要是否仅包含可从文章推断出的信息？– **属性 D：** 摘要的语法是否正确？

每篇文章的 3 个生成导语都会以随机顺序呈现给标注员，且模型名称会被隐藏。每篇文章由三名标注员进行评估，我们通过多数投票计算最终答案。

我们对每个属性—模型组合的答案取平均值，其中 1 表示“是”，0 表示“否”。图 2 展示了人工评估的结果。根据人工标注员的评估，mT5-small 模型在三个模型中表现最差。我们还观察到该模型经常出现词形变化错误。根据内容选择的质量，mT5-base 模型给出的摘要最佳，但与 Bert2Bert 模型相比，它提供的信息与文章内容不一致的情况更多。后者产生的幻觉少于另外两个模型。语法问题的平均答案显著高于其他问题，这意味着我们的模型很少生成语法错误的摘要。

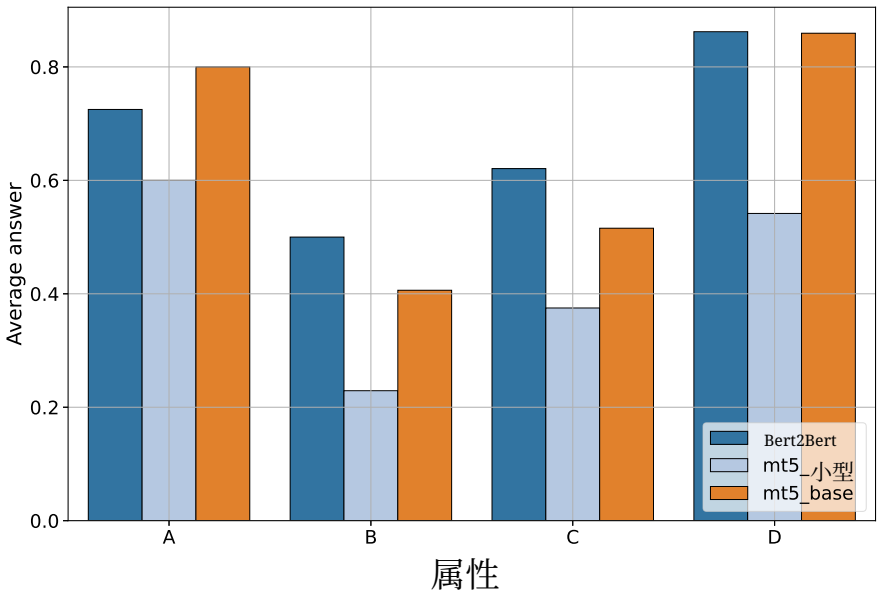

图 2：各模型在各项属性上的平均答案。

我们使用科恩κ系数（Cohen, 1960）来衡量标注员对每个问题的意见一致程度。我们计算了每一对标注员的这一数值，并取其平均值。两两标注员之间的科恩κ系数平均值为 56.77%。

语法问题的得分最低（37.99%），因为指南不够明确，一些标注员认为未完成的句子在语法上不正确，而另一些则不这么认为。

7 结论

我们介绍了*HunSum-1*，这是一个匈牙利语抽象式摘要语料库，包含从 9 家主要新闻媒体收集的 110 多万个数据样本。我们计算了该语料库质量的多项内在指标。随后，我们训练了 3 个摘要模型：一个 Bert2Bert 模型和两个 mT5 模型。我们发现，较大的 mT5 模型在自动指标和人工评估中均表现最佳。

参考文献

Agócs, Á., Yang, Z.Gy.: 使用 presumm 方法进行抽象式摘要（2022年）Bommasani, R., Cardie, C.: 摘要数据集的内在评估。载于：2020年自然语言处理实证方法会议（EMNLP）论文集。第 8075–8096 页。计算语言学协会，线上（2020年11月），https://aclanthology.org/2020. emnlp-main.649 Cheng, J., Lapata, M.: 通过抽取句子和词语进行神经摘要。CoRR abs/1603.07252（2016），http://arxiv.org/abs/1603.07252 Cohen, J.: 名义量表的一致性系数。《教育与心理测量》20(1)，37–46（1960），https://doi.org/10.1177/ 001316446002000104 Devlin, J., Chang, M.W., Lee, K., Toutanova, K.: BERT：用于语言理解的深度双向变换器预训练。载于：2019年计算语言学协会北美分会会议：人类语言技术论文集，第 1 卷（长短篇论文）。第 4171–4186 页。计算语言学协会，明尼阿波利斯，明尼苏达州（2019年6月），https://aclanthology.org/N19-1423 Feldmann, Á., Váradi, T., Hajdu, R., Indig, B., Sass, B., Makrai, M., Mittelholcz, I., Halász, D., Yang, Z.Gy.: HILBERT：云环境中匈牙利语 bert-large 模型的训练。Berend G., Gosztolya G., Vincze V.（编）《第十七届匈牙利计算语言学会议》论文集，第 29–36 页（2021年）Gehrmann, S., Deng, Y., Rush, A.: 自底向上抽象式摘要。载于：2018年自然语言处理实证方法会议论文集。第 4098–4109 页。计算语言学协会，比利时布鲁塞尔（2018年10月至11月），https://aclanthology.org/D18-1443 Hasan, T., Bhattacharjee, A., Islam, M.S., Mubasshir, K., Li, Y.F., Kang, Y.B., Rahman, M.S., Shahriyar, R.: XL-sum：面向 44 种语言的大规模多语言抽象式摘要。载于：计算语言学协会年会与自然语言处理国际联合会议 2021年论文集。第 4693–4703 页。计算语言学协会，线上（2021年8月），https://aclanthology.org/2021. findings-acl.413

Hermann, K.M., Kocisky, T., Grefenstette, E., Espeholt, L., Kay, W., Suleyman, M., Blunsom, P.: 教机器阅读和理解。载于：Cortes, C., Lawrence, N., Lee, D., Sugiyama, M., Garnett, R.（编）《神经信息处理系统进展》。第 28 卷。Curran Associates 公司（2015年） Indig, B.: 共同爬取最终也会形成语料库——从 common-crawl .hu 域名构建语料库。载于：Vincze, V.（编）《第十四届匈牙利计算语言学会议》（MSZNY 2018）。第 125–134 页。塞格德大学信息学研究所信息学系，塞格德（2018年） Joulin, A., Grave, E., Bojanowski, P., Douze, M., Jégou, H., Mikolov, T.: Fasttext.zip：压缩文本分类模型。arXiv 预印本 arXiv:1612.03651（2016a） Joulin, A., Grave, E., Bojanowski, P., Mikolov, T.: 高效文本分类的技巧集合。arXiv 预印本 arXiv:1607.01759（2016b） Lewis, M., Liu, Y., Goyal, N., Ghazvininejad, M., Mohamed, A., Levy, O., Stoy- anov, V., Zettlemoyer, L.: BART：用于自然语言生成、翻译和理解的去噪序列到序列预训练。载于：计算语言学协会第 58 届年会论文集。第 7871–7880 页。计算语言学协会，线上（2020年7月），https://aclanthology.org/2020.acl-main.703 Lin, C.Y.: ROUGE：摘要自动评估工具包。载于：《文本摘要拓展》。第 74–81 页。计算语言学协会，西班牙巴塞罗那（2004年7月），https://aclanthology.org/W04-1013 Liu, Y.: 微调 BERT 以进行抽取式摘要。CoRR abs/1903.10318（2019），http://arxiv.org/abs/1903.10318 Liu, Y., Lapata, M.: 使用预训练编码器进行文本摘要。载于：2019年自然语言处理实证方法会议暨第九届自然语言处理国际联合会议（EMNLP-IJCNLP）论文集。第 3730–3740 页。计算语言学协会，中国香港（2019年11月），https://aclanthology.org/ D19-1387 Makrai, M., Tündik, Á.M., Indig, B., Szaszák, G.: 面向匈牙利语抽象式摘要（2022年） Nallapati, R., Zhai, F., Zhou, B.: SummaRuNNer：用于文档抽取式摘要的循环神经网络序列模型。载于：第三十一届 AAAI 人工智能会议论文集。第 3075–3081 页。AAAI’17, AAAI 出版社（2017年） Narayan, S., Cohen, S.B., Lapata, M.: 别告诉我细节，只要摘要！用于极限摘要的主题感知卷积神经网络。载于：2018年自然语言处理实证方法会议论文集。第 1797–1807 页。计算语言学协会，比利时布鲁塞尔（2018年10月至11月），https://aclanthology.org/D18-1206 Nemeskey, D.M.: 语言建模的自然语言处理方法。博士论文，厄特沃什·罗兰大学（2020年） Nenkova, A., Maskey, S., Liu, Y.: 自动摘要。载于：计算语言学协会第 49 届年会论文集：

教程摘要。第3页。计算语言学协会，俄勒冈州波特兰（2011年6月），https://aclanthology.org/P11-5003 Orosz, G., Szántó, Z., Berkecz, P., Szabó, G., Farkas, R.：HuSpaCy：面向工业应用的匈牙利语自然语言处理工具包（2022）Parveen, D., Ramsl, H.M., Strube, M.：基于图的抽取式摘要的主题连贯性。载于：2015年自然语言处理实证方法会议论文集。第1949–1954页。计算语言学协会，葡萄牙里斯本（2015年9月），https://aclanthology.org/D15-1226 Paulus, R., Xiong, C., Socher, R.：用于抽象式摘要的深度强化模型。载于：第六届国际表征学习会议，ICLR 2018，加拿大不列颠哥伦比亚省温哥华，2018年4月30日至5月3日，会议专题论文集。OpenReview.net（2018），https://openreview.net/forum?id=HkAClQgA- Raffel, C., Shazeer, N., Roberts, A., Lee, K., Narang, S., Matena, M., Zhou, Y., Li, W., Liu, P.J.：探索统一文本到文本转换器的迁移学习极限（2020）Rothe, S., Narayan, S., Severyn, A.：利用预训练检查点完成序列生成任务。《计算语言学协会会刊》8，264–280（2020），https://aclanthology.org/2020.tacl-1.18 Wan, X.：探索同时进行单文档和多文档摘要的统一方法。载于：第23届国际计算语言学会议论文集（Coling 2010）。第1137–1145页（2010）Woodsend, K., Lapata, M.：故事亮点自动生成。载于：计算语言学协会第48届年会论文集。第565–574页。计算语言学协会，瑞典乌普萨拉（2010年7月），https://aclanthology.org/P10-1058 Xue, L., Constant, N., Roberts, A., Kale, M., Al-Rfou, R., Siddhant, A., Barua, A., Raffel, C.：mT5：一种大规模多语言预训练文本到文本转换器。载于：2021年北美计算语言学协会分会会议：人类语言技术论文集。第483–498页。计算语言学协会，线上（2021年6月），https://aclanthology.org/2021.naacl-main.41 Yang, Z.Gy.：以匈牙利语建模“易货”：远离世界，遥之又遥（2022）Yang, Z.Gy., Agócs, Á., Kusper, G., Váradi, T.：匈牙利语抽象式文本摘要。《数学与信息学年刊》53，299–316（2021年1月）Zhang, J., Zhao, Y., Saleh, M., Liu, P.：PEGASUS：基于抽取式缺失句的预训练，用于抽象式摘要。载于：III, H.D., Singh, A.（编）《第37届国际机器学习会议论文集》。机器学习研究论文集，第119卷，第11328–11339页。PMLR（2020年7月13日至18日），https://proceedings.mlr.press/v119/zhang20ae.html Zhou, Q., Yang, N., Wei, F., Huang, S., Zhou, M., Zhao, T.：通过联合学习句子评分与选择进行神经文档摘要。载于：计算语言学协会第56届年会论文集

语言学（第1卷：长篇论文）。第654–663页。计算语言学协会，澳大利亚墨尔本（2018年7月），https://aclanthology. org/P18-1061