

机器写作技术及应用

谢海华 助理研究员

北京雁栖湖应用数学研究院 (BIMSA)

大数据与人工智能 (BDAI)

2022.06

机器写作：机器智能生成文本内容

- 原创：计算机根据输入的结构化的数据(报表、RDF数据等)进行文字创作。该方式能够生成原创稿件，适用于[天气预报](#)、[医疗报告](#)、[赛事简讯](#)、[财经报道](#)等文本的生成。
- 二次创作：计算机根据已有文字素材(例如，已经发表的新闻)进行二次文字创作，该方式能够基于已有稿件创作出不一样的稿件，例如为一篇新闻生成[摘要](#)，对多篇相关论文进行[综述](#)，对一篇作文进行[文字改写](#)等。
- 混合创作：结合原创及二次创作两种方式进行文字创作，稿件中一部分内容从结构化数据[中直接生成](#)，另一部分内容则从已有文本中进行[提炼或改写](#)，生成的内容更丰富。

机器写作的应用

- 速报类应用。比赛信息、股讯快报等需要瞬时生成的文章。时效性高，速度快。
- 知识类应用。科普性质的文字材料，应用领域主要包括教育(诗词、历史知识)、生活、旅游等。
- 资讯聚合类应用。在现有资讯信息的基础上，重新智能聚合生成新的文章。比如，某一个电影热映，会对电影的主演、之前的作品等信息做一些扩展和盘点；对热点事件，对发展过程和关键信息进行分析聚合，形成事件脉络，便于用户了解事件全貌等。

机器写作公司 Automated Insights

- 办公智能化：为企业自动生成可读的[市场分析或收益报告](#)，赋能所有市场人员成为数据导向型的分析师（用户：Nvidia）
- 文本生成：生成数以百万计的[体育战报](#)和面向粉丝的球星[个人比赛总结](#)，从而提升用户访问时间和点击率（用户：Yahoo）
- 数据可视化产品集成：[数据的可视化](#)不仅有图表的形式还有文本，AutoIns 集成入 Tableau、MicroStrategy 等企业化数据分析和可视化软件，通过自动文本生成提升数据的可读性。

国外新闻机器写作发展历程

2001，谷歌实现个性化新闻推荐，开创了用机器选编新闻的先河；

2006，美国汤姆森公司用机器人记者撰写经济和金融方面的新闻；

2008，路透社的Open Calais 在校对界大显身手；

2011，Narrative Science 公司机器人用算法把数据转化成财经和房地产报道；

2012，《华盛顿邮报》新闻核查机器人Truth Teller；

2013，纽约公共广播的数据新闻团队制作土温传感器，准确报道了美国东岸蝉的回迁；

2014，AI 公司的机器人Wordsmith 针对读者生产定制版内容；

2014.03，《洛杉矶时报》机器人Quakebot 在地震发生三分钟后自动生成和发布了报道。除了灾难新闻，《洛杉矶时报》还开发出快速发布犯罪新闻的机器人；

2014.04，《卫报》#Open001 机器人生产出靠算法编辑的纸质报纸；

2014.07，美联社全面利用机器人Wordsmith 写作，仅需0.3 秒就可以撰写、发布上市公司盈利报道，还能定制多种语言风格；

2015，法国《世界报》和Syllabs 公司合作，用机器人记者报道了选举活动；

2015.08，《纽约时报》的机器人编辑Blossom 每天会从300多篇文章中挑出“潜力股”，推荐给编辑。其平均点击量是普通文章的38 倍；

2016.08，《华盛顿邮报》采用写稿软件Heliograf 报道里约奥运会，几秒钟即可生成并发布一条Twitter 新闻。

国内新闻机器写作发展历程

2015.09，腾讯财经推出了国内第一篇由 Dreamwriter 撰写的“机器人新闻”。

2015.11，新华社推出“快笔小新”，从事体育和经济信息报道。

2016.05，阿里巴巴与第一财经联合推出“DT 稿王”，其写稿“多、快、好”。

2016.08，《今日头条》推出 Xiaomingbot，实时撰写里约奥运会新闻稿件，截止2017年5月，Xiaomingbot 自动撰写与发布新闻共计5139 篇，包括长篇新闻420 篇，总计阅读量超过1800 万次。

2017.01，南方都市报推出的南都机器人“小南”，一秒钟就完成了春运稿件的写作。

2017.04，据新华社报道，全球首个机器人记者——中国智能机器人“佳佳”诞生，并越洋采访了美国著名科技观察家凯文·凯利。

2017.11，腾讯举办的媒体+峰会上 Dreamwriter 更是以平均0.5 秒/篇新闻的速度写出了14 篇新闻稿件，并自动配上现场图片。

2018.01，百家号推出基于人工智能的辅助写作平台“创作大脑”。创作大脑集成百度最强大的人工智能技术，是一个能够帮助作者极大提升创作效率和质量的开放创作平台。

2018.05，Giiso 写作机器人V2.0 上线，这款智能写作工具同时也是一款自媒体写作助手，擅长财经、娱乐、科技等领域的写作，创作热点文章、股评、行业简报、研究报告等文稿可做到秒级成稿。

2019.03，由齐鲁晚报倾力打造的人工智能(AI)写手“小壹”落地。

人工智能写作平台

排名	名称
1	智搜 Giiso
2	腾讯写作机器人
3	小发猫
4	神码 AI
5	张小明 (Xiaomingbot)
6	Writing-bots
7	李白写作
8	快笔小新
9	小封机器人
10	梦幻写手
11	创作大脑
12	微软小冰

机器写作技术

- 从数据到文本的生成式写作，比如：从股市数据得到的财经分析、从体育比赛数据得到的体育战报。
 - 通过经验丰富的领域媒体专家，或者机器从历史文本中挖掘，得到标准化的写作结构和句式。
 - 使用内容语料库、数据分析的结果来填充这些模板，就得到了一篇篇格式类似但言之有物的模板化新闻了。
- 从海量文本到摘要的抽取式写作，比如：搜索引擎对同一主题新闻进行摘要生成新闻主题卡片等。

地震新闻自动写作流程



四川阿坝州九寨沟县发生7.0级地震

2017-08-08 中国地震台网

速报参数

据中国地震台网正式测定，8月8日21时19分在四川阿坝州九寨沟县发生7.0级地震，震源深度20千米，震中位于北纬33.20度，东经103.82度。



震中地形

震中5公里范围内平均海拔约3827米。



热力人口

据移动人口大数据分析，震中20公里范围内人口数约2.1万，50公里范围内约6.3万，100公里范围内约30万。

周边村镇

本次地震周边5公里内的村庄有比芒，20公里内的乡镇有漳扎镇。

1
取标题

2
发布震中地图，
包括地理位置图
和地形图。

4
写地震周边的
历史情况

3
写地震参数，包括时
间、地点、震级以及
震源深度等信息

5

写地震空间分析，设定
范围（如5公里），列
出范围内的所有村庄

6

写周边的乡镇分布情
况，震源周边20公里
内的乡镇情况

8

写震中天气情况，
是否适合救援

7

列出震中县城的基本
情况，包括地理位
置、气候特征等

9

震中的人口情况，
数量多少，人员是
否集中等

10

自动发布，
无需审阅签发。



资讯配图

根据资讯的内容，匹配与资讯内容相关的图片

所属分类：写作机器人

可接入类型：API、SDK

申请服务

采集机器人 >

编辑机器人 >

写作机器人 >

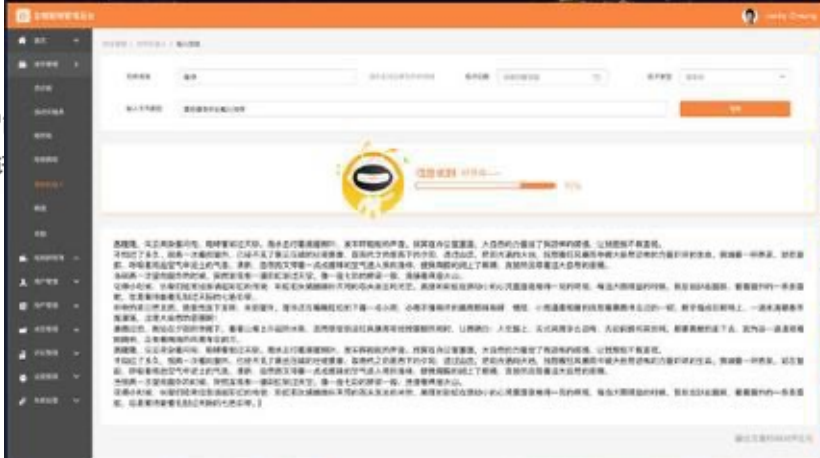
推荐机器人 >

分析机器人 >

服务介绍

通过Giiso智
可应用于资

全橙智媒



AI写作助手

基于数据、算法、逻辑监控全网

实时抓取和检索相关数据

智能学习写作风格

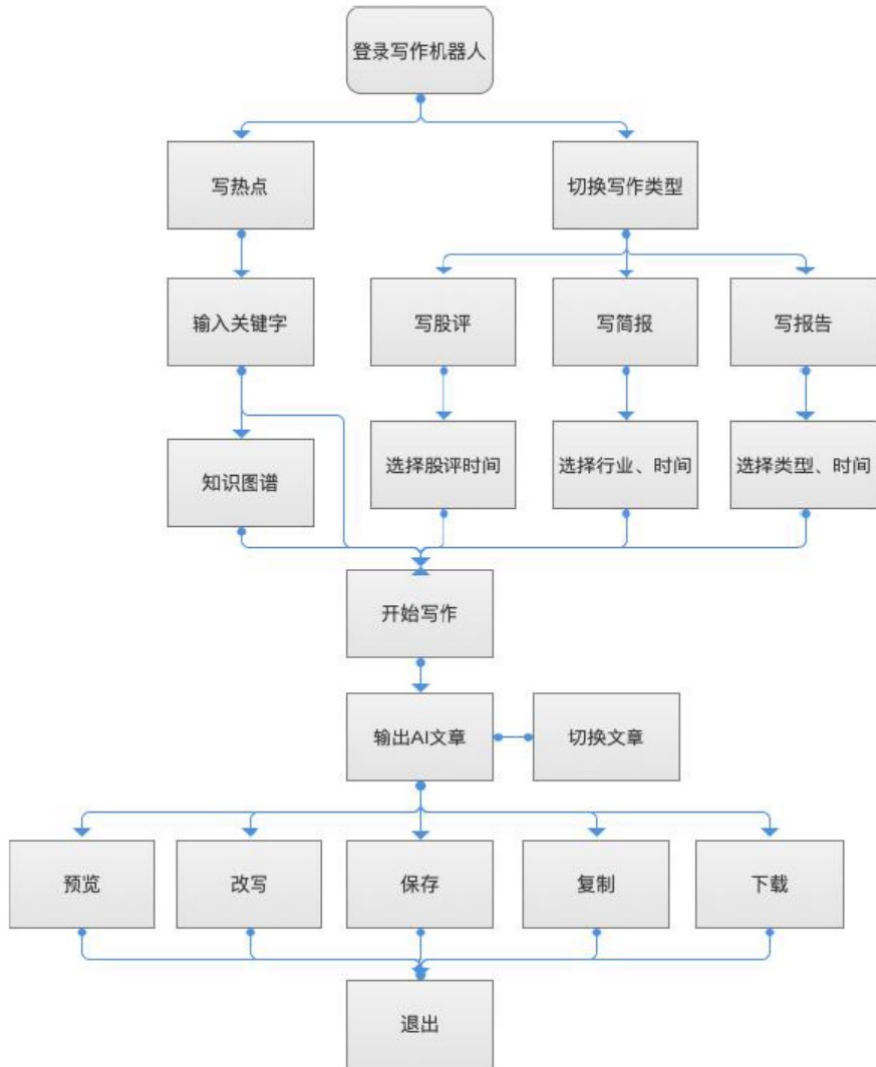
自动生成标题

智能创作文章

百家号/写作机器人

智搜 (Giiso)

- 智能语义。语义识别最高精度达到92.67%；平均精度84.22%；同类方法中最高。
- 知识图谱。基于知识图谱构建的追踪引擎最高精度95.06%；平均精度94.52%；远高于传统的搜索算法。



系统介绍

文章展示

案例展示

Dreamwriter自动写作平台，以技术创新推动内容生产革新，它重新定义了机器写作的一系列规范，并将一次写作流程抽象成通用的范式，用户可以方便的定义自己的协作规则，或利用已成熟的写作模板，来生成自己的文章内容。

系统介绍

文章展示

业务类型	文章标题	文章内容
stock	海信家电涨2.11%，创近3个月新高，报12.11元	查看详情
stock	开盘：恒指高开2.52%报23539.910点 航空航天军工板块领涨	查看详情
weather	娄底涟源05月31日天气预报	查看详情
weather	河北省邢台市发布雷电黄色预警	查看详情

海信家电涨2.11%，创近3个月新高，报12.11元

腾讯证券 2020-06-01 09:30

★ 收藏

0 评论

分享

腾讯证券讯，06月01日消息，今日早盘，海信家电（000921）涨2.11%，创近3个月新高，报12.11元，成交额566万元，换手率0.05%，振幅0.84%，量比8.33。

海信家电所在的白色家电行业，整体涨幅为0.7%。领涨股为日出东方（603366），领跌股为澳柯玛（600336）。

资金方面，海信家电近5日共流入8281.79万元，今日到此时主力资金总体呈净流出状态，净流出147.3万元。

该股最近一日（2020年05月29日）融资融券数据为：融资余额15531.51万元，融券余额123.23万股，融资买入1769.0万元，融资净买入-80.12万元。

最新的2020年一季报显示，公司实现营业收入75.86亿元，净利润4,355.04万元，每股收益0.03元，市盈率11.42。

小发猫(xiaofamao.com)



北京雁栖湖
应用数学研究院
YANQI LAKE BEIJING INSTITUTE OF
MATHEMATICAL SCIENCES AND APPLICATIONS



小发猫AI智能写作v1.1.8

[首页](#) [产品](#) [演示](#) [开发者](#) [关于](#)



加入QQ群

小朋友请在这里输入你的作文↓ 还可以输入600个字, 注册后为6000字

Large empty text area for inputting the essay.

生成AI版本

☒ 我已阅读《[小发猫服务协议](#)》并自愿遵守该协议

人工智能内容重写 AI Rewrite

原文↓

今晨，微软在纽约的秋季新品发布会上，发布了 Surface Pro 6、Laptop 2 以及最顶级的 Studio 2 三款备受期待的 Surface 产品。至此，包括年初的 Surface Book 2 在内，完成了2018 年 Surface 产品线所有升级计划。

当然这场规模不算大的发布会上，微软还带来了令人惊喜的 One More Thing，来自于全家桶套餐新添的小成员——Surface Headphones——“Surface”终于可以被听到，一款无线降噪耳机，还支持微软的语音助手 Cortana。

小发猫AI+仿写↓

全选复制

再来一篇

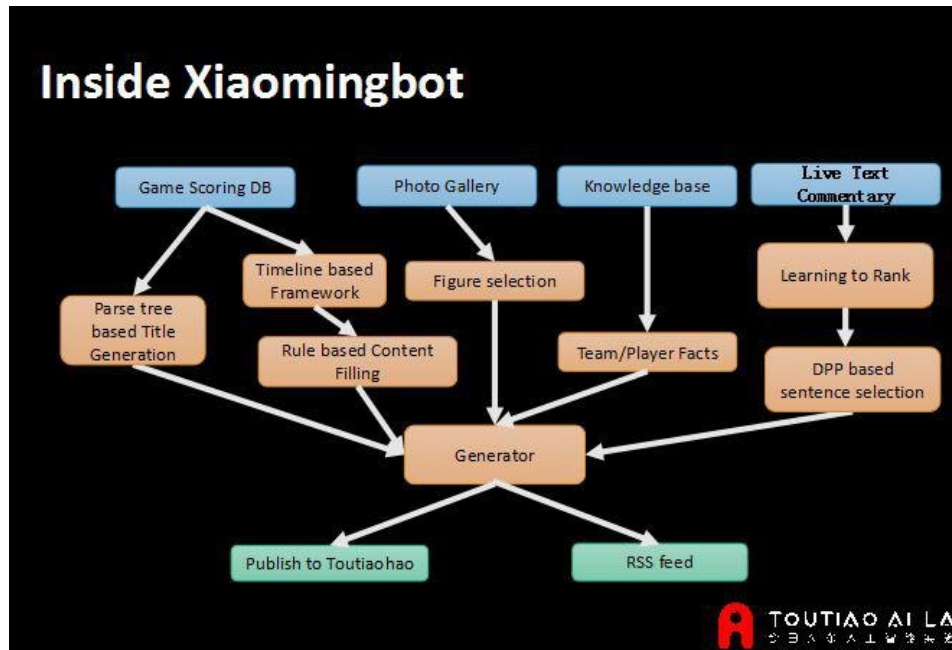
今天早上，微软在纽约秋季新产品发布会上推出了Surface Pro 6、笔记本电脑2和顶级Studio 2三款备受期待的Surface产品。到目前为止，包括今年年初的Surface Book 2，2018 Surface产品系列的所有升级计划都已完成。

当然，这不是一场大型会议，微软也带来了一个惊喜，从新系列小包装——表面耳机——“Surface”终于可以听到，一款无线降噪耳机也支持微软的语音助手Cortana。

小明Bot (Xiaomingbot)

视频生成新闻技术

- 1) 准备比赛的实时比分的数据；
- 2) 通过计算机视觉分析图片内容，将图片和文字结合；
- 3) 建立知识库，比如比赛球队的历史，球员信息等；
- 4) 抓取网络上直播文字信息，通过机器学习选取最重要的内容，要求挑选的句子相似度要尽量小，涵盖信息量尽量大，生成新闻。



小明Bot (Xiaomingbot)

摘要生成

通过摘要的方法把不同领域中一天最核心的信息摘出来，最后合成一篇文章。

1) 通过层次化的LSTM 模型对文章做建模。

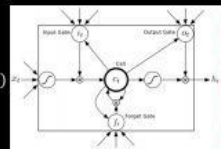
第一个维度是句子层面，通过循环神经网络对句子建模，学习句子中的语义信息，并且结合传统的模本特征，最后组合成一个向量代表每一个句子。

2) 通过LSTM 模型选择代表文章中心思想的句子，作为最后的摘要。头条每天有200 万篇的文章，通过摘要服务自动生成摘要。

Deep Summarization via Hierarchical LSTM

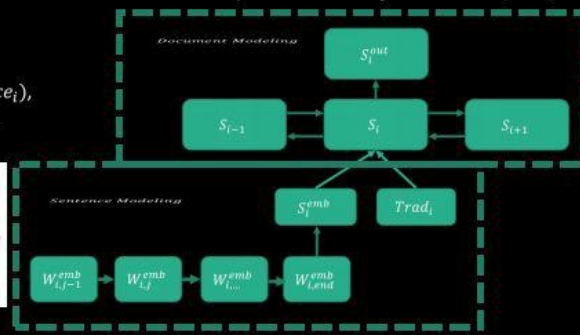
Sentence Modeling:

- Word Encoder:
 - $S_i^{emb} = biLSTM_{sent}(Word_{i,j}), j \in Sentence_i$
- $Trad_i = TraditionalFeatureExtractor(Sentence_i)$,
such as NER, TF-IDF...
- LSTM:
 - $f_t = \sigma(W_f[h_{t-1}, x_t] + b_f)$
 - $i_t = \sigma(W_i[h_{t-1}, x_t] + b_i)$
 - $\tilde{C}_t = \tanh(W_C[h_{t-1}, x_t] + b_C)$
 - $C_t = f_t * C_{t-1} + i_t * \tilde{C}_t$
 - $o_t = \sigma(W_o[h_{t-1}, x_t] + b_o)$
 - $h_t = o_t * \tanh(C_t)$



Document Modeling:

- $S_i^{out} = biLSTM_{docu}(S_i^{emb} + Trad_i), i \in Document$
- $Score_i = RankingFunction(S_i^{out})$
- $Summary = MaximumMarginalRelevance(Score)$

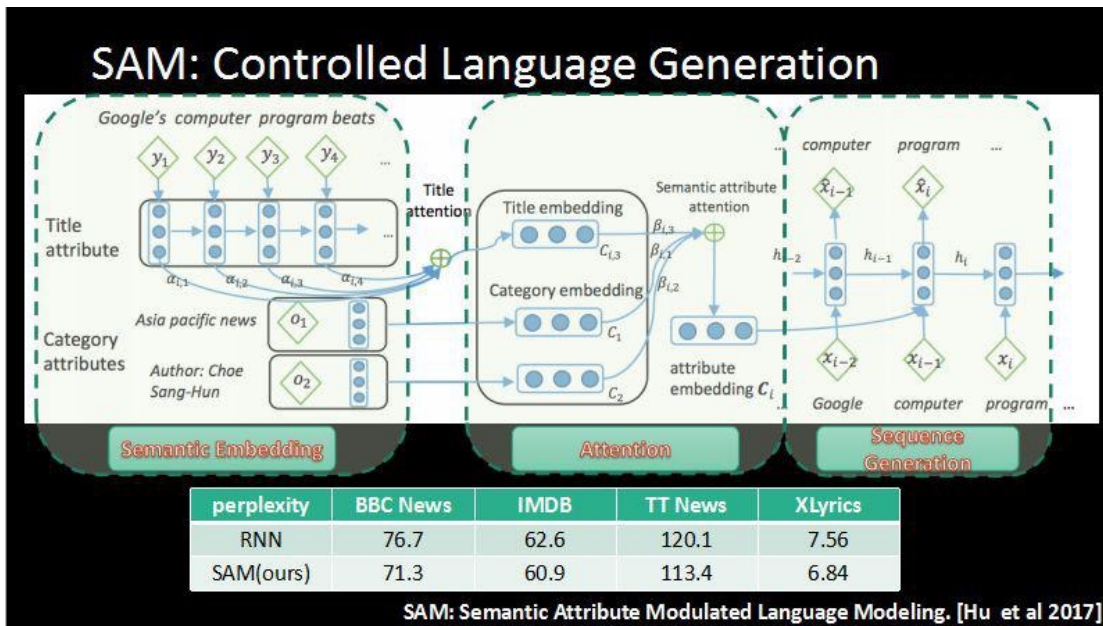


小明Bot (Xiaomingbot)

自由语言生成

通过Free Style Language Generation 模型，学习创作者的文章及写作风格，以控制风格进行写作。比如说命题作文；或者指定某一类的文章，财经类、科技等不同品类的文章，生成不同写作的风格的文章。

- 1) 通过对标题建模，对语义属性建模，把所有这些都变成语义的向量，对标题建模用了SAM；
- 2) 语义信息重要程度不同，我们通过一个机制区分这里面哪个语义标签是重要的。
- 3) 通过循环神经网络。但在这个循环神经网络里面，我们加入了前两步学到的语义信息去控制风格。这个方法我们在数量级上做测试发现越小越好，优于传统方法。



小明Bot (Xiaomingbot)

自动配图

- 1) 根据标题或者句子，通过自然语言理解技术自动地从东方IC 图库中挑选比较匹配的图片，配到文章中。
- 2) 封面图配图。为了配合封面固定大小，利用人脸识别等技术识别图片中最关键的区域，设置为封面。目前，今日头条识别技术检测时间只需要10 毫秒。

视频创作

- 1) 视频自动剪辑，比如足球比赛最关键的片段，利用自动分析技术对视频进行剪辑，把视频中最关键的信息摘出来组成片断自动播放；
- 2) 通过单一图像分析，对视频中的一帧图像分析视频中最主要的人体部位，以及各部位的位置。通过这一技术实现人体分割以及背景变换，目前已经应用到抖音小视频。

百度大脑-智能创作



请选择新建的项目类型

自动创作

自动生成不同类型的内容，快速批量且高效



新建项目

辅助创作

为您的创作提供辅助工具，撰写效率提升



新建项目

百度大脑-智能创作

自动写作支持从数据到文本描述的自动生成，即当输入源为数据和文字模板，输出为自然语言文本。

辅助写作从素材收集、文章撰写、文章检查等角度为用户提供辅助功能。

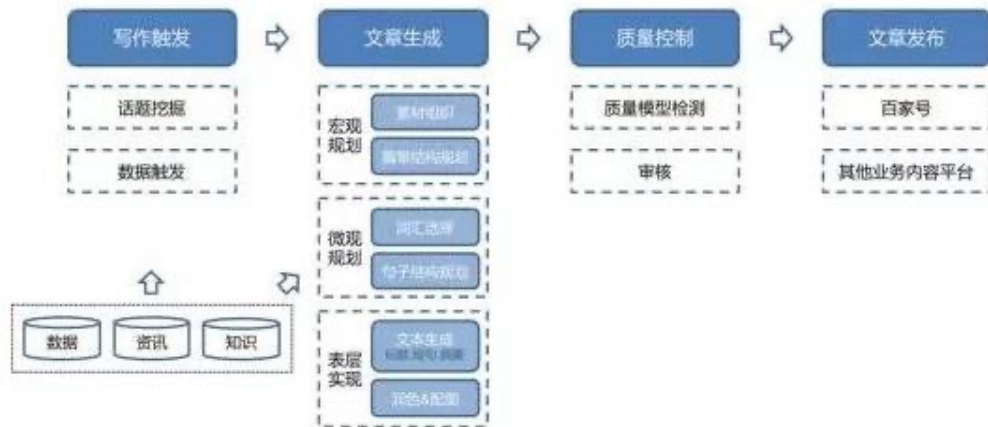


百度大脑-智能创作

自动写作技术

自动写作技术能够让机器自主的完成文章写作。当前计算机已经能够自动的撰写新闻快讯、热点组稿、春联等类型的文章。

自动写作流程

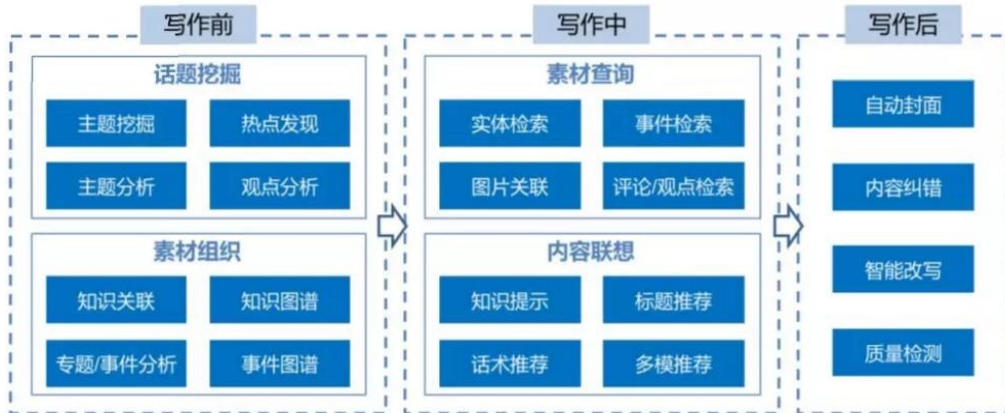


百度大脑-智能创作

辅助写作技术

提供领域热点事件发现、热点事件脉络、文本纠错和自动摘要能力，从素材收集、文章撰写、文章检查三个角度辅助创作，提升写作效率。

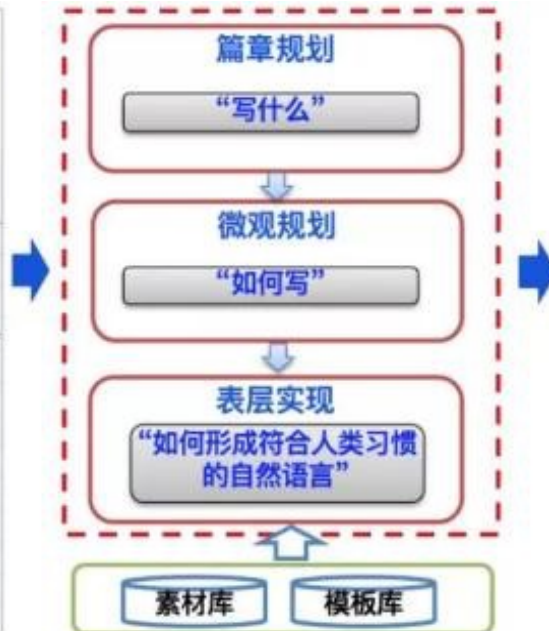
辅助写作功能



百度大脑-智能创作-技术方法

1. 经典自然语言生成算法

从篇章规划（写什么）—到微观规划（如何写）—再到表层实现（转换为自然语言）来逐步按照“流水线”进行生成算法。



百度大脑-智能创作-技术方法

2. 知识图谱

知识图谱就是知识的汇集、整理以及再加工，图谱中的每条边，均是基于语义的链接，是一个极其复杂的知识语义网络。

目前百度知识图谱数据包含亿级别实体以及千亿级别的事实，以专家权威、百科实体、垂类挖掘与全网属性挖掘为组成部分，可以做到高时效性的秒级更新，在智能写作中扮演着核心角色，贯穿智能写作的全部流程。

3. 话题挖掘

话题挖掘需要挖掘提取出用户有需要的、且适合机器写作的主题。首先从微博、feed 内容等资源中，通过知识图谱提取、匹配出热点事件与概念，接下来判断是否适合机器写作，过滤掉太发散、太泛的话题；生成的话题包括非个性化与个性化的话题，其中个性化话题是通过用户画像进行定向下发。

4. 素材组织

通过知识图谱来驱动，通过事件触发，匹配财经、市场、板块等领域素材，进一步融合理论知识与权威评论，考虑一些归纳分析等方法，得到最终的素材关联与组织形式。

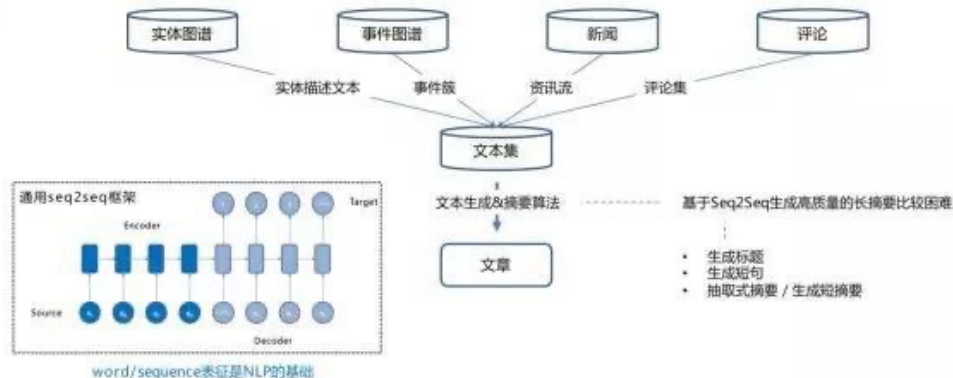
百度大脑-智能创作-技术方法

关键技术 – 文本生成



关键技术 – 文本生成

文本到文本生成 (text2text)



百度大脑-智能创作-技术方法

5. 文本生成技术

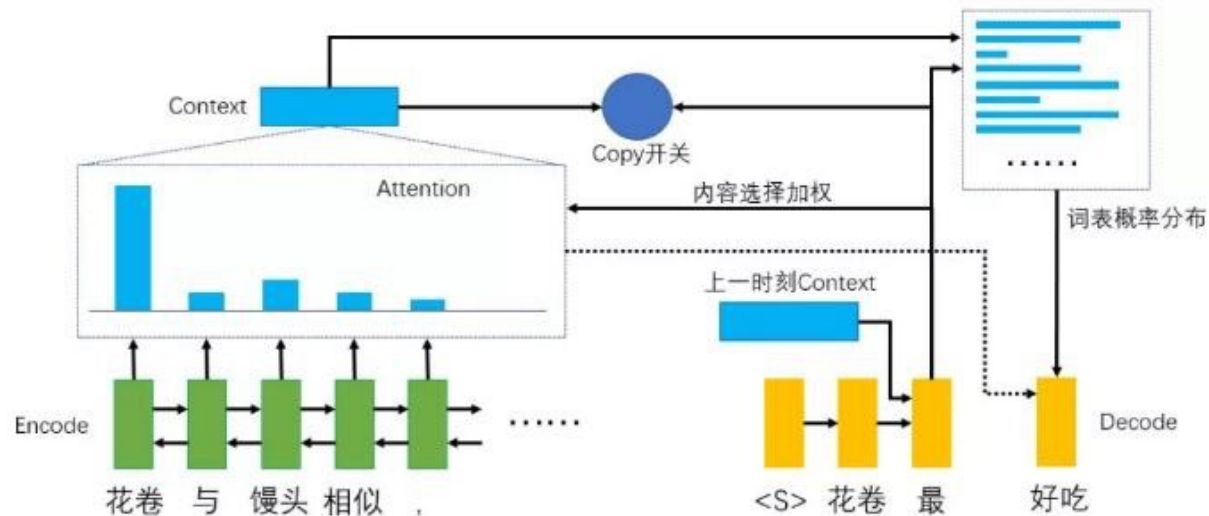
1) 摘要生成方法。在seq2seq+attention 模型的基础上引入pointer network机制构造出了新的文本摘要模型。这个模型既能够从源文本中选择复制单词，同时还保留从固定词汇集中生成单词的能力，在loss 上对重复出现的词进行打压，取得了不错的效果。

2) 从事件脉络生成聚合类文章。

例如，针对嫦娥四号发射时间，首先从事件图谱中检索相关的时间点与事件，生成相应的事件脉络。之后通过篇章规划、自动文摘，生成相关的聚合文章。

3) 标题生成，示例：

百家号创作大脑序列生成算法



花卷与馒头相似，但为卷状。花卷营养丰富，味道鲜美，做法简单。将面制成薄片拌好作料后卷成半球状，蒸熟即可。花卷的种类繁多，可根据自身爱好加入调料制作……

花卷最好吃的做法，做法简单，一上桌就被抢光！

百度大脑-智能创作-技术方法

5. 文本生成技术

4) data2text 方法。主要基于模板的方法进行生成，首先通过对现有资讯中的文本组织形式学习，通过bootstrap 算法自动生成相关的模板，再加以人工修正与设置触发条件。当有新的数据进入，则根据模板生成相应的文章。

5) 多模到文本生成方法。主要依赖于知识图谱与视频理解技术，通过视频分析，从标题、关键帧、字幕等数据源，获取相应的多模实体解析。再通过知识图谱进行关联，进行联合推断产出相关的文字。

更多的自动内容生成技术

多模态

多语言



百度大脑-智能创作

应用场景

- 1) 媒体与内容创作行业。适用于财经、体育、天气、热点与娱乐事件等多样内容的自动写作与辅助写作，大幅提升创作效率；
- 2) 商业智能。适用于企业内外部数据的自动分析与报告生成，提升企业信息同步效率与管理效率；
- 3) 行业报告与咨询机构。适用于对行业热点咨询、最新动态等信息的追踪、监测与报告的自动生成，大幅提升信息处理效率；
- 4) 市场营销与活动。运用智能写诗与智能春联的技术，帮助企业打造更多让用户有参与感的营销活动，提升营销的用户体验与传播价值。

商品介绍

清单专辑

图文结合

产品热点



乔丹 男鞋 新款跑步鞋经典复古鞋时尚休闲运动鞋 黑色/白色

立即生成文章

操作小贴士：请在下拉菜单内选择你想呈现的商品，点击“立即生成按钮”，即可看到文章样例



乔丹复古休闲鞋



李白写作

大气休闲百搭，多样颜色满足你的不同需求，缓冲减震性能优越，着地平稳，经典款式设计，时尚感十足，打造男士身形，散发着浓烈的安全与可靠性。



商品介绍

清单专辑

图文结合

产品热点



时尚智能手表，送男友的贴心礼物

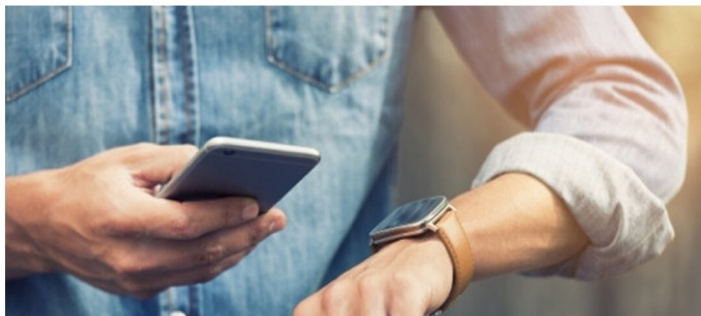


立即生成文章

操作小贴士：请在下拉菜单内选择你想呈现的商品，点击“立即生成按钮”，即可看到文章样例

时尚智能手表，送男友的贴心礼物

智能手表作为近几年兴起的新潮产品，不仅能够让我们的日常生活更加便捷，更拥多样的主题，佩戴衣着方便随心。为大家推荐以下几款智能手表，送给男友绝对会让他爱不释手哦，戴上它更帅气，更有型哦。



京东-李白写作



北京雁栖湖
应用数学研究院
YANQI LAKE BEIJING INSTITUTE OF
MATHEMATICAL SCIENCES AND APPLICATIONS

商品介绍

清单专辑

图文结合

产品热点



秦始皇兵马俑文物仿制陶瓷软木杯垫 青铜天鹅

立即生成文章

操作小贴士：请在下拉菜单内选择你想呈现的商品，点击“立即生成按钮”，即可看到文章样例

值得你种草
工艺礼品低价

立即购买



京东-李白写作



北京雁栖湖
应用数学研究院
YANQI LAKE BEIJING INSTITUTE OF
MATHEMATICAL SCIENCES AND APPLICATIONS

商品介绍

清单专辑

图文结合

产品热点



百草味 坚果炒货 休闲零食 蟹香瓜子仁158g/袋

立即生成文章

操作小贴士：请在下拉菜单内选择你想呈现的商品，点击“立即生成按钮”，即可看到文章样例



百草味 坚果炒货 休闲零食 蟹香瓜子仁158g/袋

口感醇厚

香气浓郁



新华社-快笔小新

- 1) 体育、财经类**新闻写作**。比如失业经济数据、股价报盘、人民币汇率报价、上市公司公告和财报、个股资金净流入流出、融资融券数据、中超比赛结果和积分公报、CBA 比赛结果和积分公报等，包括国外的相关数据，如英国 CPI、英国失业率情况、欧元区失业率数据、欧元区 CPI 初值以及终值数据。迅速处理，生成相应的稿件。
- 2) 运用文本摘要技术实现上**市公司公告摘要**。上市公司每天都要发布大量上市公司公告，编辑记者很难对每篇公告都进行详细阅读和跟踪。通过使用文本摘要技术，对于原文中的句子分析、评估和抽取，根据已经设计好的算法模型，自动分析和摘取其中的要点和知识点，输出短小的摘要，为编辑记者提供素材服务。
- 3) 新闻**热点分析**。快笔小新通过采集新浪微博上多位在金融领域有权威和声望的专家微博，动态关注并采集其发在微博的相关言论，对其相关内容进行关联分析，分析结果作为素材供编辑记者参考。

第一财经-财经类智能写作工具

三种文章类型

- 1) 简单的数字讲故事。根据高频数据，加上逻辑分析及预判条件，从新闻的角度，把重要的、异常的信息提取并完成新闻稿件。
- 2) 信息提取。重要信息的总结功能，比如，每天都有很多上市公司公告发布，数据量非常大而且复杂，通过信息提取的方式判断哪些公告是有价值的，然后生成摘要。
- 3) 多种信息关联。比如，根据CPI 报告，同时收集同一段时间之内的经济学家、观察家、意见领袖的看法，把关联度高的内容匹配出来，自动生成文章。

第一财经-财经类智能写作工具

技术特性：“多、快、好”

- 1) “多”方面：通过海量抓取、海量分析，主要针对上市公司公告、财务报表、官方发布、社交平台、证券行情等信息源，日阅读3000 万字，30 天可写完一部四库全书。
- 2) “快”方面：利用多台服务器分布式的对发布的公告进行扫描，可以即时对公告进行分类筛选出待写新闻稿的公告，并通过后台的信息提取算法提取出该公告的重要信息，机器写稿平均阅读7,471 字/秒= 448,275 字/分钟，是普通人阅读速度的50 倍。然后，按新闻稿成稿格式重新组织输出，机器写稿为28 字/秒= 1,680 字/分钟，普通人打字的35 倍。
- 3) “好”方面：生成的稿件语句流利通顺，完全符合人类自然语言的语法。解决了目前机器人写稿的最大问题，即语句不通顺不流畅。在海量的信息中抓住重点信息本身的重点，协助决策的快速实施。如东旭光电科技股份有限公司5 月14 日所发布的公告为例，全文2,274 个字，总结成113字的核心内容和23 字的标题，同时不是盲目地抓取公告中的大标题。



笔神作文

用科技提升学习效率

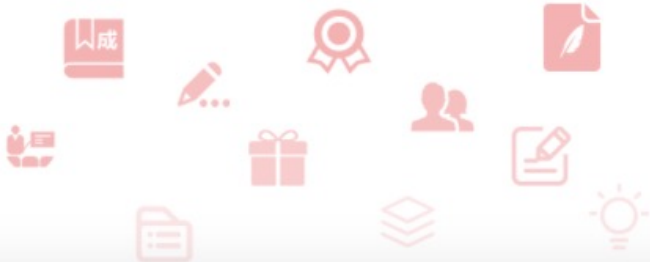
人工智能作文学习工具

精选范文 智能讲文 联想描写

小学

初中

高中



高中

搜索作文

全部 叙事 写人 写景 议论 状物 游

叶落禾木

转眼又到了一年的秋天，一叶知秋，城市里早熟的树叶已开始发黄，露出一丝秋意，想想此刻的禾木该是秋意正浓。

山中日月长，每想起那年秋天在喀纳斯禾木乡住的那一日，就像是昨天才发生的事。那令人心醉的美，让禾木秋日里的时光在心里停留至今。

禾木被称为喀纳斯的后花园，神的自留地。我们赶到禾木时已近黄昏，木板搭建的农舍、几根圆木棍横平竖直围起的简易牛圈、依旧青绿的草场，自然、安详、

高一 叙事 700字

笔神评分：89分

智能推荐描写

人工智能分析所有范文的段/句/词

基于NLP技术拆解范文，提供更多描写方式

转眼又到了一年的秋天，一
叶知秋，城市里早熟的树叶已开
始发黄，露出一丝秋意，想想此
刻的禾木该是秋意正浓。

秋天

树叶

城市

秋天，迷人的秋景令人神往；
秋天，蓝天、白云、红叶相互
映衬；一处处秋景，一幅幅秋
画，一堆堆果实，尽情渲染着
秋意浓浓，令人心旷神怡，美
不胜收！

珍视友谊

初雨的缘—友谊

友谊

友谊在我们的心中好似一座桥。如果这座桥坍塌，而我们的友谊也随之烟消云散了。友谊是一幅无色的画，值得我们去增添色彩。友谊是一道美味的菜肴值得我们去慢慢品味。

友谊是一本博大精深的书本，值得我们去领悟里面的知识韵味！

让我们敞开胸怀去描绘灿烂的友谊吧

高三 散文 1500字

笔神评分：94

范文匹配

基于作文内容进行自动化分类匹配

用户仅通过左右滑屏即可轻松浏览内容相关范文

范文

素材

草稿

不管是顺境也好、逆境也好，不管自己处在何种境遇，都要抱着积极的心态朝前看，任何时候都要拼命工作，持续努力，这才是最重要的。 - 稻盛和夫

冬天越寒冷，樱花就开得越烂漫。

不要同情自己，同情自己是卑劣懦夫干的勾当。
——《挪威的森林》

世上有可以挽回的和不可挽回的事，而时间经过就是一种不可挽回的事。 - 《国境以南太阳以西》村上春树

每个人都有属于自己的一片森林，也许我们从来不曾去过，但它一直在那里，总会在那里。迷失的人迷失了，相逢的人会再相逢。 ——村上春

学习素材库

范文、推荐描写词句一键收藏

基于用户阅读范文内容，推荐学习匹配的美词佳句

Q 高铁

热门主题: [不负韶华](#) [归园田居](#) [三潭映月](#) [花港观鱼](#) [雷锋夕照](#) [双峰插赤](#) [断桥残雪](#) [题西林壁](#) [夜雨寄北](#) [不惧风雨](#)

搜索Topk:

- 2 +

作诗耗时: 58351n

AI作诗结果

天险何年凿
山腰盘地脉
马去沙无迹
谁能将铁笛

Q 北京

热门主题: [不负韶华](#) [归园田居](#) [三潭映月](#) [花港观鱼](#) [雷锋夕照](#) [双峰插赤](#) [断桥残雪](#) [题西林壁](#) [夜雨寄北](#) [不惧风雨](#)

搜索Topk:

- 2 +

作诗耗时: 303ms

AI作诗结果

万国来朝会,
天威清朔漠,
玉帛通华夏,
中原今一统,

Q 回乡

热门主题: [不负韶华](#) [归园田居](#) [三潭映月](#) [花港观鱼](#) [雷锋夕照](#) [双峰插赤](#) [断桥残雪](#) [题西林壁](#) [夜雨寄北](#) [不惧风雨](#)

搜索Topk:

- 2 +

作诗耗时: 96

AI作诗结果

十年踪迹滞
白发满头犹
风尘荏苒双
莫怪临歧频

Q 西安

热门主题: [不负韶华](#) [归园田居](#) [三潭映月](#) [花港观鱼](#) [雷锋夕照](#) [双峰插赤](#) [断桥残雪](#) [题西林壁](#) [夜雨寄北](#) [不惧风雨](#)

搜索Topk:

- 2 +

作诗耗时: 1n

AI作诗结果

秦城西去接
王气不随天
金汤百二雄
回首夕阳频

五言律诗

Q 敦煌

热门主题: [不负韶华](#) [归园田居](#) [三潭映月](#) [花港观鱼](#) [雷锋夕照](#) [双峰插赤](#) [断桥残雪](#) [题西林壁](#) [夜雨寄北](#) [不惧风雨](#)

搜索Topk:

- 2 +

作诗耗时: 12

AI作诗结果

汉家天子爱
今日中原无

Q 作诗

热门主题: [不负韶华](#) [归园田居](#) [三潭映月](#) [花港观鱼](#) [雷锋夕照](#) [双峰插赤](#) [断桥残雪](#) [题西林壁](#) [夜雨寄北](#) [不惧风雨](#)

搜索Topk:

- 2 +

作诗耗时: 624ms

AI作诗结果

作诗

学诗不得吟哦趣, 作赋徒劳感慨多。
自是胸中有丘壑, 敢将文字问如何。

备注:
1、调节不同的搜索Topk可以生成不同的古诗;
2、访问人数较多, AI作诗可能会有延迟, 请耐心等待。

机器写诗-技术

三个主要步骤：

- 1) 第一步：让机器学习约50万首律诗和绝句，形成几个能表现“题目”，“句子”和“词汇”三者关系的数据库。
- 2) 第二步：根据题目，使用与题目相关的词汇形成每个句子的第一个词汇，然后再形成每个句子，然后机器先合成出一首原型诗。
- 3) 第三步：即时根据格律要求、词汇间关系的紧密程度，优化原型诗，直到符合格律或达到设定的优化次数，最后输出。

机器写诗-技术

步骤一：机器学习了什么？

以《望庐山瀑布》为例：

日照香炉生紫烟，遥看瀑布挂前川。飞流直下三千尺，疑是银河落九天。

1) 首先是分词：

(a) 题目分词为：望/庐山/瀑布。其中庐山和瀑布是专用名词，所以这首诗的主题可能是“庐山”或“瀑布”。

到底是“庐山”还是“瀑布”？需要如下分析：

取出这首诗内容的词汇，与其他诗词进行匹配。基于一定的匹配度获取相关诗词，并分析相关诗词的题目。如果相关诗词的题目中含有“瀑布”更多，则认为此诗的主题是“瀑布”（否则相反）。

(b) 诗的内容分词（2-2-2-1或者2-2-1-2）

例诗内容分词结果：

日照/香炉/生/紫烟，遥看/瀑布/挂/前川。

飞流/直下/三千/尺，疑是/银河/落/九天。

机器写诗-技术

2) 分词之后, 形成三种数据库:

- 推荐词数据库: 词汇之间前后关系的数据库。
- 共现词库: 表现题目和词汇关系的数据库。
- 对词库: 对仗词数据库。

(a) 推荐词数据库

什么是“推荐词”? 例诗里: “日照”是“香炉”的前推荐词, “香炉”是“日照”的后推荐词。推荐词数据库不只标识词汇间的前后关系, 也标识词汇间“有没有关系”。

(b) 共现词库

什么是共现词? 同一主题下, 共同出现的词汇。例诗的主题是“瀑布”。将数十万首诗词进行分类, 比如以“瀑布”为主题的诗词有几百首。这几百首诗里经常共同出现的词汇称为“瀑布”的共现词(例如: 峰前、银河、飞泉、清谷、壁山、烟霞.....)

(c) 对仗词库

分词之后, 得到对仗词(约100万组不重复的对词), 例如: “天上”和“人间”, “落去”和“归来”。

机器写诗-技术



北京雁栖湖
应用数学研究院
YANQI LAKE BEIJING INSTITUTE OF
MATHEMATICAL SCIENCES AND APPLICATIONS

瀑布

[瀑布]两字前推荐词

千寻	遥看	奔腾	来看	香炉	半天
不同	窗喷	窗前	倒悬	飞空	飞来
飞泉	风含	风前	峰高	扶看	高成
何如	恍疑	更多			

[瀑布]三字后推荐词

独逡巡	半空落	冰成日	不飞津	出空微
穿危石	穿云落	当公署	登高阁	滴铜瓶
涤烦襟	斗直泻	飞灵雨	飞晴雪	飞晴雨
飞素练	分僧榻	古来闻	挂峰头	挂江北

瀑布

[瀑布]词的热度为:98%(在诗词 20万,常用词中排名第 3513)

[瀑布]和该词走得最近的词

庐山	香炉	白云	千寻	青山	匡庐
人间	五老	烟霞	明月	天台	遥看
银河	不知	登临	飞流	峰前	何处
梅花	声中	更多			

[瀑布]对仗词

香炉	梅花	庐山	山屏	柏台	碑文
苍松	草堂	禅徒	菖蒲	更多	

机器写诗-技术

步骤二：根据题目，合成一首原型诗

- 怎么根据题目获取与题目相关的词汇？
- 怎么根据词汇形成句子？
- 律诗怎么处理？

1) 基于用户给定的题目获取词汇：从共现词库中获取。

例如：如果用户给出的题目有“瀑布”，则共现词可以有：飞流、银河、飞泉、清谷、壁山、烟霞..... 选出其中四个分配给四个句子。

2) 根据词汇形成语句：从前后推荐词库里获取，以瀑布为例：

从前后推荐词库中，查找瀑布，可以看到：

前推荐词有：千寻、奔腾、倒悬、飞来、窗前.....

后推荐词有：半空落、出空微、穿危石、挂峰头、飞晴雨.....

把“前推荐词”“瀑布”“后推荐词”进行组合：

千寻瀑布半空落，飞来瀑布出空微。倒悬瀑布挂峰头，窗前瀑布飞晴雨。

3) 律诗的合成。绝句中间加上两联，形成律诗（基于对词库进行构造）。五律：将七律的前两个字去除。

机器写诗-技术

步骤三：优化和输出

从四个共现词（绝句为例）中，以格律为约束条件，通过交叉和迭代，收敛到同一个韵部。

为了保证趣味性，每次都要形成不同的诗词，因此不能优化到最佳值（只有一个）。算法只优化到符合格律（平仄、押韵）为止，把更多的选择以候选句的形式提供给用户。

题目 天涯 体裁 七绝

创作 校验 配图

千里迢迢返故庐
平生寂寞故人疏
江山不见莺花老
落日相逢岁又除

请选择诗句替换 确认

梅花落日辞青琐
江山不见莺花老
当年日月无消息
青山不是安家好

机器写作的优势

1. 使新闻采写更高效；
2. 减少报道差错；
3. 对海量数据的再生产；
4. 语言样式接近传统记者写作风格。

机器写作的问题

冗余信息的威胁

过去社会迫切需要解决的问题是如何获得更多的信息，而如今新媒介已使世界成为一个信息垃圾堆放场，对信息的追求使人们淹没在信息的汪洋大海里。“**为了对付新的信息，就需要增补控制机制。但新的控制机制本身就是技术，它们又反过来增加信息的供应量。**”写作机器人诞生的初衷之一就是依靠其信息处理效率，将读者从海量信息中解放出来。但其本身的高产不可避免地进一步推动信息爆炸的趋势，流动规模也将更加庞大。这种“以亿计”的新闻报道以及一些因超量制作的文稿形成的大量信息垃圾，可能会带来一定的负面效应，让“信息焦虑症”变得更为严重。

机器写作的问题

是否解放了人力

即使在内容生产环节解放了人力，在后期的审核、编辑环节人力的分量是不是并未减少甚至提高了呢？比如腾讯 Dreamwriter 团队基本不会插手后期审核的工作，同时也不会对机器人的稿件进行加工润色。但是根据一些国外机器人写作的经验发现，有的媒体会对机器人写作的产品进行人工审核或编辑，例如美联社。哥伦比亚大学新闻学院导师 Tom Kent 曾在文章中说：“**机器和数据出错时不可避免的，但一旦出错可能会导致成千上万的新闻稿件发生错误，因此在进行写作之前需要彻底对其进行审查，在稿件生成之后，也需要编辑在发布之前对其进行核查。**即是说，我们所看到的机器人写出的作品都或多或少带有人力审查的痕迹。”

机器写作的问题

写作伦理问题

若新闻机器人写出了虚假新闻等不合规范的新闻，该负责的是机器还是背后的人？例如 2014 年 3 月，美国《纽约邮报》报道，失踪的 NBA 球员坤顿·罗斯尸体已被找到，此新闻被其他新闻机构迅速抓取并转发，其中就有谷歌开发的“维基百科实时检测机器人” (Wikipedia Live Monitor)。12 分钟后，《纽约邮报》发表更正说，弄错了，不是坤顿·罗斯，而是一个同名同姓的死者，但机器人因设计逻辑局限(多人同时修改维基百科的同一条目才启动信息抓取和发布程序)并未重新抓取和更正信息，导致原来的误报流传。算法可以快速地采集和发布新闻，也可以快速犯错。因其高效率和人员干涉成本高，一旦犯错负面影响较大。若是权威的信息源，新闻出错可能性小；但如果其信源来自社交媒体，操控的空间比较大。再者，机器人写作在向图片新闻、视频新闻发展，如何去抓取真实的、不侵犯他人权利、不伤害他人的图片或视频，这个边界目前还没有完善的解决办法。

问题及讨论

Q&A