

源 APIExp 使用手册

——源模型 API 图形化工具使用手册

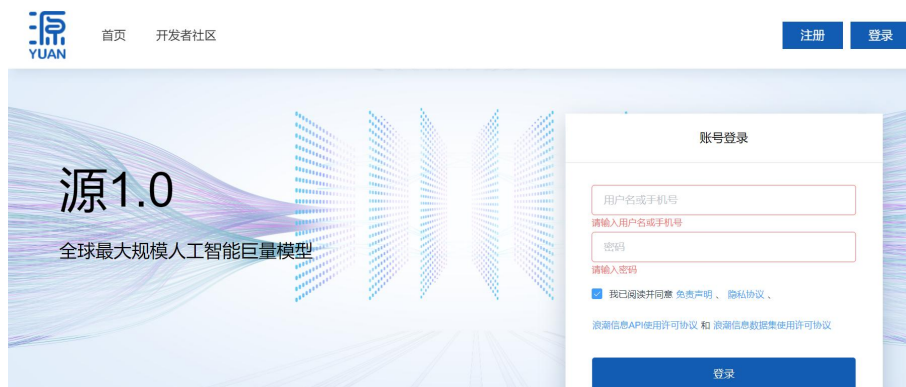
文档时间：2022-05-11

1 源网站登录

打开源网站（<https://air.inspur.com>），注意：源网站的网址输入正确，https 写成 http，将会导致打开页面失败。点击右上角【登录】按钮，进入登录页面。



在登录页面，输入注册的用户名或手机号、密码进行登录。如果忘记了密码，请联系源服务邮箱 air_service@inspur.com 进行密码重置。



2 控制台进入

API 申请已经通过审核获得授权的用户，登陆后在右上角会出现【控制台】。

如果没有出现【控制台】可能是因为①还没有提交使用 API 的申请或者②申请还未通过。①未提交申请的开发者，可以点击【**申请使用 API**】按钮（下图红色部分），填写申请表进申请。②已提交申请，申请还未通过审核，请耐心等待，一般在 5 个工作日内会获得申请的结果，如要加急可发邮件到 air_service@inspur.com 或者联系源工作人员微信 [joejeanjean](#)。



点击【控制台】按钮，进入到用户的控制台界面，在左侧菜单栏可以看到有三个菜单，[API 余量]，[APIExp]和[操作日志]。默认展示 API 余量页面，点击[APIExp]或[操作日志]菜单，可以进入到对应的页面。



注意：在 API 余量页面，可以看到用户授权的信息。如果授权已过期，即使用次数用完或者无剩余天数，可以邮箱或微信联系源工作人员申请授权延期。

3 使用 APIExp

APIExp 是源模型 API 服务的图形化操作工具。在图形化的 APIExp 页面，可以方便的使用源模型的 API 推理服务，选择所使用的源模型，调整模型对应的超参数，不需要撰写任何代码，就可以使用源模型的能力，来实践自己的创意想法。



3.1 模型选择

目前源模型 API 推理服务上线了 3 款预训练语言模型，参数规模都是百亿级别，分别是基础模型和两款领域模型（对话模型和翻译模型）。基础模型是通用的语言模型，具备的能力更广泛，通常在专有领域的效果不如领域模型。不过这也不是绝对的，你可以多加体验，来感受不同模型的能力。



3.2 超参数配置

超参数你可以选择源模型提供的默认参数，不需要有任何改动，这些默认参数是源算法工程师多次测试后选取模型效果还不错的一组超参数组合。当然，你可以调整这些超参数，来探究更好的模型使用效果。

3.2.1 超参数说明

目前超参数一共有 7 个，详细说明如下表。

参数名	参数作用	取值范围
最大返回长度	模型推理返回结果的最大 token 序列长度。该参数设置越大，模型生成答案耗时越长。设置过短可能影响生成结果的完整性。token 可能是一个字、词或者标点。	[1-200]
temperature	temperature 值越大，模型的创造性越强，但生成效果不稳定。temperature 值越小，模型的稳定性越强，生成效果稳定。	(0-1]
topP	生成 token 的概率累加，从最大概率的 token 往下开始取，当取到累加值大于等于 topP 时停止。当 topP 为 0 时，该参数不起作用。	[0-1]
topK	挑选概率最高的 k 个 token 作为候选集。若 k 值为 1，则答案唯一。当 topK 为 0 时，该参数不起作用。	[0-200]
输入前缀	输入文字的开头符号，如“问题是：”	[0-8]个字
输出前缀	输出文字的开头符号，如“答案是：”	[0-8]个字
输出停止符	模型输出的停止符。生成文本遇到该字符停止，截断输出。如设定为“。”，则模型生成的文本结果从第一个“。”进行截断。	单个符号

3.2.2 超参数默认值

参数名	“基础模型”默认值	“对话模型”默认值	“翻译模型”默认值
最大返回长度	50	50	50
temperature	1.0	1.0	1.0
topP	0.8	0.7	0.8
topK	5	5	5
输入前缀	空	空	空
输出前缀	空	空	空
输出停止符	空	空	空

3.3 输入示例

源 1.0 模型是语言模型，输入一段文本，模型就会输出一段文本。在输入区进行文本输入，输入完成点击【提交】按钮，就将输入区的文本提交给了源模型，

耐心等待输出即可。

源 1.0 预训练语言模型是支持 zero-shot 和 few-shot 使用的，zero-shot 和 few-shot 的意思是给模型一些“例子”作为先验知识，zero-shot 是指不给任何例子，few-shot 是指给出一些例子。

3.3.1 zero-shot 使用

在输入区进行文本输入，输入完成点击【提交】按钮，就将输入区的文本提交给了源模型，耐心等待输出即可。比如，输入：用“首屈一指”这个词造句吧。

The screenshot shows a web interface for APIExp. At the top left is the APIExp logo. Below it is a text input area containing the prompt "用“首屈一指”这个词造句吧。". At the bottom of the input area is a blue button with the text "提交" (Submit).

点击提交后，返回结果为：“我是首屈一指的！”

The screenshot shows the same APIExp interface as before, but now the output "“我是首屈一指的！”" is displayed below the input prompt. The "提交" (Submit) button remains at the bottom.

3.3.2 few-shot 使用

如果完全不给例子使用效果不好，就可以先给出几个例子。比如：用“故园”为题做首诗吧。模型给出的结果看起来是以“故园”为题在写小作文，而不是一首诗。

APIExp

以“故园”为题做首诗。
“故园”的含义，在不同的人心中有不同的理解。对于一个人而言，故乡是他的出生地；对于一个家庭而言，故乡是祖祖辈辈生活的地方；对于一个民族

提交

这是因为预训练语言模型是一种通用模型，当使用在特定任务时，它并不一定理解“做首诗”具体的涵义。给出一个例子让模型理解“做首诗”的涵义，就可以看到模型输出的结果好了很多。只是这次模型是依靠“记忆”给出的，直接给的是贺知章的《回乡偶书》诗的内容作为结果。（这是不是有点像人考试一样，能靠记忆的绝不靠思考 $o(*\text{---}*\text{---}*)o$)

APIExp

以“风”为题做首诗：解落三秋叶，能开二月花。过江千尺浪，入竹万竿斜。
以“故园”为题做首诗：
少小离家老大回，乡音无改鬓毛衰。儿童相见不相识，笑问客从何处来。

提交

3.3.3 prompt 优化


为了让模型输出更好的结果，我们还可以给模型“提示”（prompt），不同的 prompt 也会影响模型的使用效果。（这是不是像，领导给下属布置工作，说不清楚工作目标和边界的话，下属可能完全往错误的方向去做~）

prompt 就是用高效的“固定表达模板”来降低交互理解的难度，以期望模型能够在“模板”规定的框框内正确地行动，得到满意的结果。

比如，**情感分类的任务**：给定一个句子，判断句子所表达的情感是高兴还是悲伤。如果**输入**：“今天终于通过了业界最难的考试”，这句话表达的情感是高

兴还是悲伤？

模型输出：为什么？（3）你认为这个故事中的人物有哪些优点？（4）你认为这个故事中的人物有哪些缺点？（5）你觉得这个故事中的人物

“今天终于通过了业界最难的考试”，这句话表达的情感是高兴还是悲伤？
为什么？（3）你认为这个故事中的人物有哪些优点？（4）你认为这个故事中的人物有哪些缺点？（5）你觉得这个故事中的人物

提交

这是因为模型认为[这个输入]是在针对“今天终于通过了业界最难的考试”出题，所以模型继续进行了问题的续写，又给补充了几个问题。

没有给出“指导”的话，模型虽然也能够返回通顺的句子，但是往往答非所问、做非所求。我们进行 prompt 优化，变换一下说法，输入：今天终于通过了业界最难的考试，他的心情真

模型输出：好！“我要把这个好消息告诉我的朋友们，让他们也分享我的快乐！”他兴冲冲地跑到街上，拦住一辆出租车，向朋友家驶去。“嗨

今天终于通过了业界最难的考试，他的心情真好！“我要把这个好消息告诉我的朋友们，让他们也分享我的快乐！”他兴冲冲地跑到街上，拦住一辆出租车，向朋友家驶去。“嗨

提交

结果看上去会好很多，只不过还不是我们要求的只输出情感标签“高兴”或“悲伤”。为了让模型能够匹配我们的要求来进行输出，这时我们就可以再给模型增加几个 few-shot 示例。

输入：今天和男朋友吵了一架，他的心情真悲伤。今天终于通过了业界最难的考试，他的心情真高兴。今天他被雨淋感冒了。他的心情真

模型输出：悲伤

（tips：模型也可能输出“难过”、“糟糕”等词，还可以调整最长返回长度、输出停止符等超参数）



The screenshot shows the APIExp web interface. At the top left is the APIExp logo. Below it is a text input field containing the prompt: "今天和男朋友吵了一架，他的心情真悲伤。今天终于通过了业界最难的考试，他的心情真高兴。今天他被雨淋感冒了，他的心情真悲伤。". Below the input field is a blue button with the text "提交" (Submit).

合适的 prompt 和高质量的 few-shot 示例，可以零代码使用源 APIExp，实践自己的创意想法，开发有趣的 AI 应用。快去尝试一下吧~