

http://

应用商店

系统管理

Hugging Face

全球大型的开源社区，专注于NLP技术

Models

52,624

search

tasks

[image Classification](#)

[translation](#)

[Fill Mask](#)

[QA](#)

[sentence summarization](#)

[text generator](#)

+ 24 tasks

libraries

[tensorflow](#)

[pytorch](#)

[JAX](#)

[transformers](#)

[ONNX](#)

+ 4

license

[Apache](#)

[MIT](#)

[GPL-2](#)

+ 30

在线商店

Hugging Face

OpenMMLab

PyTorch Hub

Tensorflow Hub

distilgpt2

★★★★☆

text generation

详情

distilgpt2

★★★★☆

text generation

详情

distilgpt2

★★★★☆

text generation

详情

distilgpt2

★★★★☆

text generation

详情

distilgpt2

★★★★☆

text generation

详情

distilgpt2

★★★★☆

text generation

详情

distilgpt2

★★★★☆

text generation

详情

distilgpt2

★★★★☆

text generation

详情

distilgpt2

★★★★☆

text generation

详情

distilgpt2

★★★★☆

text generation

详情

distilgpt2

★★★★☆

text generation

详情

distilgpt2

★★★★☆

text generation

详情

跳转到 1 页 1/12723 < >

http://

应用商店

系统管理

OpenMMLab

深度学习时代最完整的计算机视觉开源算法体系

在线商店

OpenMMLab

OpenMMLab

PyTorch Hub

Tensorflow Hub

Models 23

search

tasks

[Rotated object detection](#)

[Human Pose and Shape Estimation](#)

[Unsupervised Learning](#)

[+ 5 tasks](#)

libraries

[tensorflow](#)

[pytorch](#)

[JAX](#)

[transformers](#)

[ONNX](#)

[+ 4](#)

license

[Apache](#)

[MIT](#)

[GPL-2](#)

[+ 30](#)

MMRotate

★★★★☆

text generation

详情

MMSelfSup

★★★★☆

text generation

详情

MMRotate

★★★★☆

text generation

详情

MMSelfSup

★★★★☆

text generation

详情

MMRazor

★★★★☆

text generation

详情

MMFlow

★★★★☆

text generation

详情

MMRazor

★★★★☆

text generation

详情

MMFlow

★★★★☆

text generation

详情

MMHuman3D

★★★★☆

text generation

详情

MMFewShot

★★★★☆

text generation

详情

MMHuman3D

★★★★☆

text generation

详情

MMFewShot

★★★★☆

text generation

详情

A Web Page

http://

KubeGems 模型商店

应用商店 系统管理

Hugging Face 全球大型的开源社区，专注于NLP技术

编辑 部署

distilgpt2

version

v1.1.0

v1.2.1

用户评分

3.5

102评价

 平台推荐

62

类型

[image Classification](#) | [translation](#)

库

[tensorflow](#) | [pytorch](#)

协议

Apache v2

发布

已发布

介绍 文件 讨论区 运行实例 证书

BERT base model (uncased)

Pretrained model on English language using a masked language modeling (MLM) objective. It was introduced in this paper and first released in this repository. This model is uncased: it does not make a difference between english and English.

Disclaimer: The team releasing BERT did not write a model card for this model so this model card has been written by the Hugging Face team.

Model description

BERT is a transformers model pretrained on a large corpus of English data in a self-supervised fashion. This means it was pretrained on the raw texts only, with no humans labelling them in any way (which is why it can use lots of publicly available data) with an automatic process to generate inputs and labels from those texts. More precisely, it was pretrained with two objectives:

Masked language modeling (MLM): taking a sentence, the model randomly masks 15% of the words in the input then run the entire masked sentence through the model and has to predict the masked words. This is different from traditional recurrent neural networks (RNNs) that usually see the words one after the other, or from autoregressive models like GPT which internally mask the future tokens. It allows the model to learn a bidirectional representation of the sentence.

Next sentence prediction (NSP): the models concatenates two masked sentences as inputs during pretraining. Sometimes they correspond to sentences that were next to each other in the original text, sometimes not. The model then has to predict if the two sentences were following each other or not.

This way, the model learns an inner representation of the English language that can then be used to extract features useful for downstream tasks: if you have a dataset of labeled sentences for instance, you can train a standard classifier using the features produced by the BERT model as inputs.

Intended uses & limitations

You can use the raw model for either masked language modeling or next sentence prediction, but it's mostly intended to be fine-tuned on a downstream task. See the model hub to look for fine-tuned versions on a task that interests you.

Note that this model is primarily aimed at being fine-tuned on tasks that use the whole sentence (potentially masked) to make decisions, such as sequence classification, token classification or question answering. For tasks such as text generation you should look at model like GPT2.


A Web Page

http://

☰ KubeGems 模型商店


应用商店

系统管理

 **Hugging Face**

全球大型的开源社区，专注于NLP技术

编辑 部署

 **distilgpt2**

version

v1.1.0

v1.2.1

用户评分

3.5

102评价

平台推荐 62

类型

image Classification

translation

库

tensorflow

pytorch

协议

Apache v2

发布

已发布

介绍 文件 讨论区 运行实例 证书

2021-06-29 10:31:23 CST 10KiB 103fe01eab40e883:17a5563be73:17a559b4ef5:4a81a981

2021-06-29 10:37:42 CST 10KiB 103fe01eab40e883:17a556a08d0:17a55a10e24:80e6db66

2021-06-29 11:32:23 CST 8.5KiB 103fe01eab40e883:17a559b89cb:17a55d2ea45:a3e5eb73

2021-06-29 11:38:42 CST 8.6KiB 103fe01eab40e883:17a55a183db:17a55d8e8bd:6dbb5324

2021-06-29 12:32:53 CST 9.3KiB 103fe01eab40e883:17a55d33921:17a560a3c53:cac0b1ae

2021-06-29 12:39:42 CST 9.2KiB 103fe01eab40e883:17a55d94abf:17a5610ac39:4528ec5d

2021-06-29 13:33:53 CST 8.0KiB 103fe01eab40e883:17a560ac5c7:17a56423fb7:8385b86d

2021-06-29 13:40:42 CST 9.6KiB 103fe01eab40e883:17a56114971:17a56489c96:a9d13a6e

2021-06-29 14:34:53 CST 9.1KiB 103fe01eab40e883:17a5642f0e1:17a567a1947:8fa3182a

2021-06-29 14:41:12 CST 7.3KiB 103fe01eab40e883:17a5648b02c:17a567fedc2:4fa486a7

2021-06-29 15:35:23 CST 11KiB 103fe01eab40e883:17a567a67d2:17a56b18fcb:d8a84a9e

2021-06-29 15:41:42 CST 13KiB 103fe01eab40e883:17a56803c1b:17a56b73b62:5e42c985

2021-06-29 16:36:23 CST 14KiB 103fe01eab40e883:17a56b20545:17a56e958f8:830b60a1

2021-06-29 16:42:12 CST 12KiB 103fe01eab40e883:17a56b7761a:17a56eeb85a:12881168

2021-06-29 17:36:53 CST 12KiB 103fe01eab40e883:17a56e98037:17a5720d6a4:608b55b7

2021-06-29 17:43:12 CST 12KiB 103fe01eab40e883:17a56ef41a6:17a5726973c:4b01234d

2021-06-29 18:37:23 CST 9.5KiB 103fe01eab40e883:17a5720ea3f:17a57580c30:5de0fd52

2021-06-29 18:43:42 CST 9.6KiB 103fe01eab40e883:17a5726f96e:17a575e0687:596bed

2021-06-29 19:37:53 CST 7.7KiB 103fe01eab40e883:17a57585ac3:17a578f9abd:918528c2

2021-06-29 19:44:12 CST 7.4KiB 103fe01eab40e883:17a575e1a19:17a57950a49:8b97fa6e

2021-06-29 20:38:23 CST 8.4KiB 103fe01eab40e883:17a578fd57a:17a57c70193:8fb03214

2021-06-29 20:45:42 CST 8.6KiB 103fe01eab40e883:17a5795ceaa:17a57cd4995:33f122a9

2021-06-29 21:38:53 CST 7.8KiB 103fe01eab40e883:17a57c71525:17a57fe52c5:5e4d2f17

2021-06-29 21:46:42 CST 7.7KiB 103fe01eab40e883:17a57cdf9d5:17a580573ad:f81f28c6

2021-06-29 22:39:23 CST 9.8KiB 103fe01eab40e883:17a57fe79f1:17a5835b564:c7d20375

2021-06-29 22:47:12 CST 10KiB 103fe01eab40e883:17a5805ae65:17a583cad8f:8aed897c

2021-06-29 23:40:23 CST 5.5KiB 103fe01eab40e883:17a5835dc88:17a586d7b47:a7c39e6c

2021-06-29 23:47:42 CST 5.4KiB 103fe01eab40e883:17a583d0f8e:17a58744d72:8bbc013c

2021-06-30 00:41:23 CST 11KiB 103fe01eab40e883:17a586e1807:17a58a5708a:fa70a68

2021-06-30 00:48:12 CST 11KiB 103fe01eab40e883:17a58748844:17a58eb7d08:2c255f95

A Web Page

http://

KubeGems 模型商店

应用商店

系统管理

Hugging Face

全球大型的开源社区，专注于NLP技术

编辑

部署

distilgpt2

version

v1.1.0

v1.2.1

用户评分

3.5

★★★★☆

102评价

平台推荐

62

类型

image Classification

translation

库

tensorflow

pytorch

协议

Apache v2

发布

已发布

介绍

文件

讨论区

运行实例

证书

评价

LinkMaq

14 days ago

★★★★☆

...

This is a nice job !

编辑

删除

10 回复

回复展开

LinkMaq

14 days ago

★★★★☆

...

This is a nice job !

10 回复

LinkMaq

14 days ago

★★★★☆

...

This is a nice job !

10 回复

A Web Page

http://

KubeGems 模型商店

应用商店

系统管理

Hugging Face

全球大型的开源社区，专注于NLP技术

distilgpt2

version

v1.1.0

v1.2.1

用户评分

3.5

★★★★☆

102评价

平台推荐

62

类型

[image Classification](#)

[translation](#)

库

[tensorflow](#)

[pytorch](#)

协议

Apache v2

状态

已上架

获取所有集群的modeldeploy cr 资源，模型运行实例不做租户权限隐藏

编辑 部署

介绍 文件 讨论区 运行实例 证书

部署实例	模型版本	状态	集群	命名空间	创建人	API	
distigpt2	v1.1.0	运行中	demo1	demo1	link.ma	https://distigpt-xgdsg.models.kubegems.io	详情
distigpt2	v1.2.1	运行中	demo2	demo2	link.ma	https://distigpt-fdstw.models.kubegems.io	详情

进入模型实例的详情页，对当前用户权限鉴权

A Web Page

⏮️ ⏪️ ❌ 🏠

http://

🔍

☰ KubeGems 模型商店

应用商店

系统管理

Hugging Face 全球大型的开源社区，专注于NLP技术

distilgpt2

version

v1.1.0

▼

v1.2.1

用户评分

3.5

★★★★☆

102评价

平台推荐

62

类型

[image Classification](#)

[translation](#)

库

[tensorflow](#)

[pytorch](#)

协议

Apache v2

状态

已上架 ✓

下一步

❗

基本配置

⚙️

详细配置

✓

完成

应用名称

distigpt2

项目

项目空间 ▼

版本

v1.0.0

▼

v1.0.1

环境

环境空间 ▼

A Web Page

← → ✕ 🏠

http://

🔍

☰ KubeGems 模型商店

应用商店 系统管理 用户头像 通知

Hugging Face 全球大型的开源社区，专注于NLP技术

📦 distilgpt2

version

v1.1.0 v1.2.1

用户评分 3.5

★★★★☆

 102评价 平台推荐 62

类型 [image Classification](#) | [translation](#)

库 [tensorflow](#) | [pytorch](#)

协议 Apache v2

状态 已上架 ✓

上一步 部署

! ⚙️ ✓

基本配置 详细配置 完成

Type

seldondeployment

protocol

seldon tensorflow v2 interface

? 什么是 protocol?

runtime

huggingface tensorflow Serving mlflow Server triton interface (ONNX,Pytorch,Tensorflow,TensorRT) SKLearn Server Tempo Server XGBoost Server

资源部署

实例数 1 分配 CPU 1 核 分配 内存 4 GB

☒ 分配 GPU

nvidia.com/gpu tencent.com/vcuda

根据目标环境上 kubegems 的 GPU 插件启动状态配置

GPU 核心 可使用 / 已分配 0.3 / 4 GPU 显存 (GB) 可使用 / 已分配 2 / 16

分配显卡 0.3 分配显存 4

vcuda 显卡最小调整单位 0.1 nvidia 显卡最小为整数

vcuda 显存最小调整单位 256MB nvidia 不做显存配置

A Web Page

⏮️ ⏪️ ❌ 🏠

http://

🔍


☰ KubeGems 模型商店


应用商店

系统管理

👤

🔔

 **Hugging Face** 全球大型的开源社区，专注于NLP技术

 distilgpt2

version

v1.1.0

▼

v1.2.1

▲

用户评分

3.5

★★★★☆

102评价

平台推荐

62

类型

image Classification

|

translation

库

tensorflow

|

pytorch

协议

Apache v2

状态

已上架 ✓

上一步 部署

❗

基本配置

⚙️

详细配置

✓

完成

🟢

部署成功 [点击查看](#)

跳转到实例详情页

A Web Page

http://

KubeGems 模型商店

应用商店

系统管理

GemsHub 深度学习时代最完整的计算机视觉开源算法体系

在线商店

OpenMMLab

OpenMMLab

PyTorch Hub

Tensorflow Hub

Models 10

search

tasks

Rotated object detection

Human Pose and Shape Estimation

Unsupervise Learning

+ 5 tasks

libraries

tensorflow

pytorch

JAX

transformers

ONNX

+ 4

license

Apache

MIT

GPL-2

+ 30

MMRotate

★★★★☆

text generation

详情

MMRazor

★★★★☆

text generation

详情

MMHuman3D

★★★★☆

text generation

详情

MMHuman3D

★★★★☆

text generation

详情

A Web Page

http://

KubeGems 模型商店

应用商店

系统管理

Hugging Face

全球大型的开源社区，专注于NLP技术

distilgpt2

version

v1.1.0

v12.1

用户评分

3.5

102评价

平台推荐

62

类型

image Classification

translation

库

tensorflow

pytorch

协议

Apache v2

状态

已上架

基本配置

详细配置

完成

Type

modeldeployment

seldondeployment

镜像地址

docker.io/sgsaga:test

镜像私钥

如何创建镜像
下载凭证

运行命令

环境变量

端口号

2389

高级配置

模型挂载路径

/models

需保证该路径下为空文件夹

访问设置

网关

default-gateway

访问域名

https://distigpt2-gsg.kubegemes.io

不填由平台根据默认网关
domain 创建随机域名

资源部署

实例数

1

分配 CPU

1

核分配 内存

4

GB

分配 GPU

nvidia.com/gpu

tencent.com/vcuda

GPU 核心

可使用 / 已分配

0.3 / 4

GPU 显存 (GB)

可使用 / 已分配

2 / 16

分配显卡

1

分配显存

4

上一步

部署

A Web Page

http://

☰ KubeGems 工作台

应用商店

系统管理

算法详情 / gpt2-test

模型信息

项目空间

demo

项目环境

demo

服务类型

modeldeployment

模型

gpt2

版本

latest

来源

huggingface

框架

pytorch

类型

text

创建人

link.ma

创建时间

2022/07/01 12:34:02

基本信息

监控

调整副本

yaml

编辑

删除

已分配 CPU

1 核

已分配 内存

16GB

已分配 GPU

0.2 GPU

已分配 显存

8GB

访问地址

API 地址

https://distigpt2-xw3sdg.kubegems.io

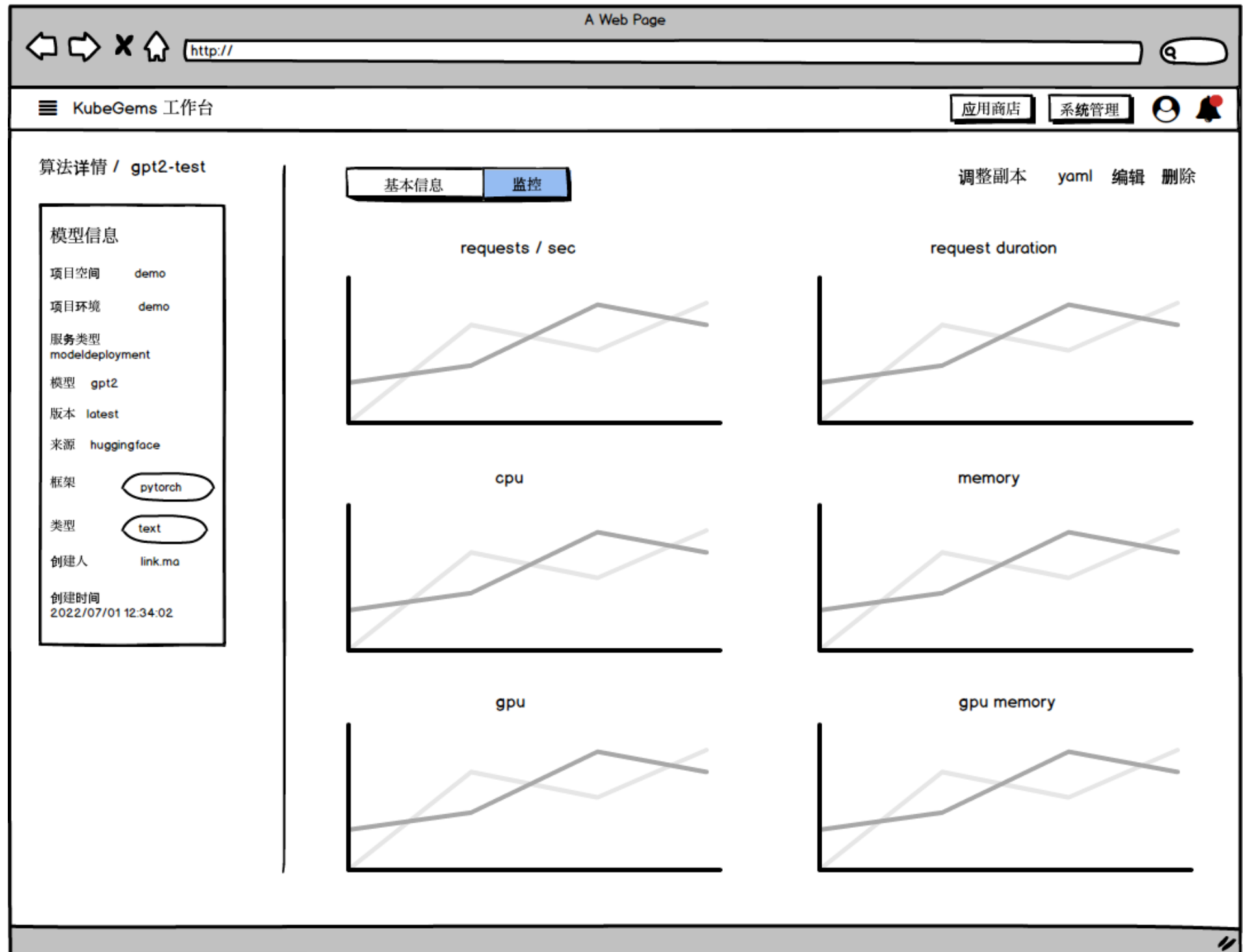
API reference

https://distigpt2-xw3sdg.kubegems.io/ap1/doc/

内部服务	端口
10.96.24.35	8000
10.96.24.35	8080
10.96.24.35	9000
10.96.24.35	9001

实例信息

容器组	状态	重启次数	age	pod ip	node	镜像版本	操作
distigpt2	Running	1h	1.	1.1.1.1	node1	sdsgsg/sgg:latest	实时日志 shell 事件 重启



A Web Page

http://

KubeGems 模型商店

应用商店

系统管理

Hugging Face

全球大型的开源社区，专注于NLP技术

部署

删除

distilgpt2

version 1.2.1

用户评分

3.5

102评价

平台推荐

62

类型

image Classification

translation

库

tensorflow

pytorch

协议

Apache v2

状态

已上架

基本信息

集群: demo

项目: demo

环境: demo

监控

已分配 CPU

1 核

已分配 内存

16GB

已分配 GPU

0.2 GPU

已分配 显存

8GB

访问地址

API 地址

https://distigpt2-xw3sdg.kubegems.io

内部地址

10.96.24.35

端口

2424

实例信息

容器组	状态	重启次数	age	pod ip	node	镜像版本	操作
distigpt2	Running	1h	1.	1.1.1.1	node1	sdsgsg/sgg:latest	实时日志shell事件重启

http://

应用商店

系统管理

Hugging Face

全球大型的开源社区，专注于NLP技术

distilgpt2

version 1.2.1

用户评分

3.5

102评价

平台推荐

62

类型

image Classification

translation

库

tensorflow

pytorch

协议

Apache v2

状态

已上架

介绍

文件

讨论区

运行实例

证书

Lincense 授权

有效期至 2022/30/12 23:59:59

sgasgsagjkls gasgjkdsjgksald,gdshguiehtbdjnkjsadgsdgas

编辑

清除

开源仓库 功能栏中没有证书， gemshub模型仓库开启

对于 gemshub 托管的模型，如 lincense 授权失败则部署按钮置灰

A Web Page

⬅️ ➡️ ✕ 🏠


http://


🔍

☰ KubeGems 管理员

应用商店

系统管理





基本配置

空间

成员

审计



集成

认证

应用商店

模型商店

prometheus

 **Hugging Face** <https://huggingface.co>  同步成功 2022/2/2 00:00:00 内置

简介：全球大型的开源社区，专注于NLP技术

Models

52,624

Image



12

同步

激活

编辑

删除

 **OpenMM Lab** <https://models.cloudminds.com>  同步失败 2022/2/2 00:00:00 内置

简介：深度学习时代最完整的计算机视觉开源算法体系

Models

23

Image


2

同步

激活

编辑

删除

 新加模型商店

←

→

✕

🏠

A Web Page

http://

🔍

☰ KubeGems 管理员

应用商店 系统管理

👤 🔔

基本配置

空间

成员

审计

集成

认证

应用商店

模型商店

prometheus

商店名称

AlGems

🔗 下载 modelx client

📄 下载 modelx client 管理模型

AWS S3 IDOE DFT OSS Azure GCS TOS

Endpoint

https://s3.kubegems.io

Region

default

🔴

skip insecure

Bucket

aigmes

AccessKey

xxxxxxxx

SecreKey

xxxxxxxx

verify

验证 S3 协议可用性的同时，校验对象存储 bucket 的根内是否存在 index.modelx.yaml 索引文件

商店管理

用户

管理员

🔍 search

👤 LinkMaq

👤 LinkMaq

👤 LinkMaq

👤 LinkMaq

🔍 search

👤 LinkMaq

👤 LinkMaq

👤 LinkMaq

取消

保存

prometheus

更新索引

```
recomment > 80,  
为模型创建一个 hot  
的 log
```