pipeline 工具函数

情感分类

```
from transformers import pipeline

classifier = pipeline('sentiment-analysis')
classifier("I'm studying AI technology and to be more successful.")
```

[{'label': 'POSITIVE', 'score': 0.9954908490180969}]

> 输出结果显示：No model was supplied，那么 pipeline 函数会选择一个特定的预训练的情感分类模型，这里选择的是 distilbert-base-uncased-finetuned-sst-2-english，而且这个模型已经对英文情感分类进行了微调。

> 第一次运行 pipeline 时会自动下载选择的模型并保存在本地默认的路径下，下次再运行相同的代码则会直接使用本地保存的模型，无需再下载。

> 还可以对多个文本进行情感分类

```
texts = [
    "I'm studying AI technology and to be more successful.",
    "I don't like my present job", "I like computer programming."
]

classifier(texts)
```

[{'label': 'POSITIVE', 'score': 0.9954908490180969}, {'label': 'NEGATIVE', 'score': 0.9987264275550842}, {'label': 'POSITIVE', 'score': 0.9976498484611511}]

零样本分类 zero-shot-classification

```
from transformers import pipeline

classifier = pipeline('zero-shot-classification')
classifier('The stock market is doing very well today',
           candidate_labels=['education', 'politics', 'business'])
```

{'sequence': 'The stock market is doing very well today', 'labels': ['business', 'education', 'politics'], 'scores': [0.8373449444770813, 0.08651848882436752, 0.07613656669855118]}

> pipeline 默认选择的模型是 facebook/bart-large-mnli，zero-shot 表示不需要对模型进行微调即可直接使用。返回的结果是文本属于给定的几个标签（由 candidate 数组定义）的概率。

文本生成 text-generation

```
from transformers import pipeline

generator = pipeline('text-generation')
generator('This course is teaching you how to')
```

[{'generated_text': "This course is teaching you how to develop a working memory. If you don't know how to code, learn using Python for now. It will be a fun and challenging one.\n\nPractical Reading\n\nThe final course I'm going to"}]

> 还可以使用参数 num_return_sequences 控制生成多少个不同的序列，使用参数 max_length 控制输出文本的总长度

```
generator('This course is teaching you how to',
          num_return_sequences=2,
          max_length=40)
```

[{'generated_text': "This course is teaching you how to design, build, and fabricate your custom designs with ease. As you get better, you'll also learn how to create and process custom projects.\n\nFor"}, {'generated_text': 'This course is teaching you how to use tools to create custom web pages based on your current interests.\n\nThis course is intended for students who do not already have a web editor, but instead want'}]

> 还可以在 pipeline 函数里通过 model 参数指定 huggingface 的其他模型

```
# 这里不使用 gpt2 模型，而是使用 distilgpt2 模型完成文本生成的工作
generator = pipeline('text-generation', model='distilgpt2')
generator('This course is teaching you how to',
          num_return_sequences=2,
          max_length=40)
```

[{'generated_text': 'This course is teaching you how to work in the computer science classroom with good practice, teamwork, and mastery.'}, {'generated_text': "This course is teaching you how to put simple, yet powerful, strategies together to make your training work for the world.\n\n\n\nIn this course dissertation, he'll start by focusing on the"}]

填充空缺 fill-mask

```
from transformers import pipeline

unmasker = pipeline('fill-mask')
masked_text = 'Python programming language is used for <mask>.'
unmasker(masked_text, top_k=3)
```

[{'score': 0.09702698886394501, 'token': 45465, 'token_str': ' scripting', 'sequence': 'Python programming language is used for scripting.'}, {'score': 0.07898325473070145, 'token': 47021, 'token_str': ' debugging', 'sequence': 'Python programming language is used for debugging.'}, {'score': 0.029713667929172516, 'token': 38228, 'token_str': ' visualization', 'sequence': 'Python programming language is used for visualization.'}]

> top_k 参数用于控制显示的结果又多少种。需要填充的部分用 mask 和尖括号进行标记，这个标记称为掩码标记，模型将会填充掩码标记所在位置的内容。

命名实体识别 ner

```
from transformers import pipeline

text = 'My name is Frank, I am a programmer, I work at Microsoft in Seattle.'

ner = pipeline('ner', grouped_entities=True)
ner(text)
```

[{'entity_group': 'PER', 'score': 0.99898106, 'word': 'Frank', 'start': 11, 'end': 16}, {'entity_group': 'ORG', 'score': 0.99948275, 'word': 'Microsoft', 'start': 47, 'end': 56}, {'entity_group': 'LOC', 'score': 0.9976749, 'word': 'Seattle', 'start': 60, 'end': 67}]

> 参数 grouped_entities=True，那么程序将会把同一类型的实体划分到同一组里面。

问答系统 question-answering

进行问答操作时要注意，不能简单的给 pipeline 一个问题，还要给它传一段完整的上下文。那么 pipeline 加载的模型将会根据问题，在上下文里找到答案。

```
from transformers import pipeline

context = "Amazon's global headquarters is located in Seattle, Washington, USA.
Alibaba is headquartered at 969 Wenyi West Road, Yuhang District, Hangzhou City,
Zhejiang Province, China"

question_1 = 'where is Amazon?'
question_2 = 'where is Alibaba?'

qa = pipeline('question-answering')
a1 = qa(question=question_1, context=context)
a2 = qa(question=question_2, context=context)

print(a1)
print(a2)
```

{'score': 0.9378065466880798, 'start': 43, 'end': 67, 'answer': 'Seattle, Washington, USA'} {'score': 0.15057657659053802, 'start': 97, 'end': 116, 'answer': '969 Wenyi West Road'}

文本摘要 summarization

```
from transformers import pipeline

# 这里为了控制视频的长度，对初始文本的内容做了删减。大家在实际操作的时候尽量保留文本的全部内容
text = '''
America has changed dramatically during recent years. Not only has the number of
graduates in traditional engineering disciplines such as mechanical, civil,
electrical, chemical, and aeronautical engineering declined, but in most of the
premier American universities engineering curricula now concentrate on and encourage
largely the study of engineering science. As a result, there are declining offerings
in engineering subjects dealing with infrastructure, the environment, and related
issues, and greater concentration on high technology subjects, largely supporting
increasingly complex scientific developments. While the latter is important, it should
not be at the expense of more traditional engineering.

Rapidly developing economies such as China and India, as well as other industrial
countries in Europe and Asia, continue to encourage and advance the teaching of
engineering. Both China and India, respectively, graduate six and eight times as many
traditional engineers as does the United States. Other industrial countries at minimum
maintain their output, while America suffers an increasingly serious decline in the
number of engineering graduates and a lack of well-educated engineers.
'''

summarizer = pipeline('summarization')
summarizer(text)
```

[{'summary_text': ' America suffers an increasingly serious decline in the number of engineering graduates and a lack of well-educated engineers . China and India graduate six and eight times as many traditional engineers as does the U.S. Other industrial countries at minimum maintain their output, while America suffers a decline in engineering graduates .'}]

> 可以指定 max_length 和 min_length 控制摘要结果的长度

```
summarizer = pipeline('summarization', min_length=30, max_length=100)
summarizer(text)
```

[{'summary_text': ' America suffers an increasingly serious decline in the number of engineering graduates and a lack of well-educated engineers . China and India graduate six and eight times as many traditional engineers as does the U.S.'}]

翻译 translation

翻译的情况有一点点特殊，因为程序并不知道人类具体的翻译要求，所以要在 pipeline 里指定模型，实现将一种语言翻译成另外一种语言。

```
from transformers import pipeline

translator = pipeline('translation', model='Helsinki-NLP/opus-mt-en-zh')
translator('I am studying Artificial intelligence')
```

[{'translation_text': '我在研究人工智能'}]

> 同样，可以指定翻译后输出文本的 min_length 和 max_length，这两个参数显然是针对于翻译较长的文本时使用的。

```
text = '''
A friend is not just someone to whom you say "Hello". A friend is a tender shoulder on which you can softly cry. A friend is a well into which you can pour all your troubles down. A friend is a hand which pulls you up from darkness and despair when all your other "so-called friends" have helped to put you there. A true friend is an ally who can't be moved or bought, a voice which keeps your name alive when others have forgot. But most of all a friend is a heart, a strong and sturdy wall, for from the hearts of friends there comes the greatest love of all!
'''

translator = pipeline('translation', model='Helsinki-NLP/opus-mt-en-zh')
translator(text, min_length=30, max_length=200)
```

[{'translation_text': '朋友不仅仅是你所说的朋友。朋友是一个温柔的肩膀,你可以轻轻地哭泣。朋友是一个井,你可以倾注你所有的烦恼。朋友是一个使你脱离黑暗和绝望的手,当你的所有其他"所谓的朋友"帮助你在那里的时候。一个真正的朋友是一个不能移动或买入的盟友,一个在别人忘记你的时候能保留你名字的声音。但大部分朋友是一颗心,一堵坚固的墙,那是从朋友的心中发出的最深的爱!'}]

总结

> pipeline 主要用于演示或者快速使用的目的，它用于完成特定的目的，不能根据细分的领域对它进行自定义修改。

> 以下再次列出 pipeline 针对特定任务的参数名称

> sentiment-analysis 情感分类

> zero-shot-classification 零样本分类

> text-generation 文本生成

> fill-mask 填充空缺

> ner 命名实体识别

> question-answering 问答

> summarization 文本摘要

> translation 翻译