第四章 文本概括

在繁忙的信息时代，小明是一名热心的开发者，面临着海量的文本信息处理的挑战。他需要通过研究无数的文献资料来为他的项目找到关键的信息，但是时间却远远不够。在他焦头烂额之际，他发现了大型语言模型（LLM）的文本摘要功能。

这个功能对小明来说如同灯塔一样，照亮了他处理信息海洋的道路。LLM 的强大能力在于它可以将复杂的文本信息简化，提炼出关键的观点，这对于他来说无疑是巨大的帮助。他不再需要花费大量的时间去阅读所有的文档，只需要用 LLM 将它们概括，就可以快速获取到他所需要的信息。

通过编程调用 AP I接口，小明成功实现了这个文本摘要的功能。他感叹道："这简直就像一道魔法，将无尽的信息海洋变成了清晰的信息源泉。"小明的经历，展现了LLM文本摘要功能的巨大优势：**节省时间**，**提高效率**，以及**精准获取信息**。这就是我们本章要介绍的内容，让我们一起来探索如何利用编程和调用API接口，掌握这个强大的工具。

一、单一文本概括

以商品评论的总结任务为例：对于电商平台来说，网站上往往存在着海量的商品评论，这些评论反映了所有客户的想法。如果我们拥有一个工具去概括这些海量、冗长的评论，便能够快速地浏览更多评论，洞悉客户的偏好，从而指导平台与商家提供更优质的服务。

接下来我们提供一段在线商品评价作为示例，可能来自于一个在线购物平台，例如亚马逊、淘宝、京东等。评价者为一款熊猫公仔进行了点评，评价内容包括商品的质量、大小、价格和物流速度等因素，以及他的女儿对该商品的喜爱程度。

```
prod_review = """
这个熊猫公仔是我给女儿的生日礼物，她很喜欢，去哪都带着。
公仔很软，超级可爱，面部表情也很和善。但是相比于价钱来说，
它有点小，我感觉在别的地方用同样的价钱能买到更大的。
快递比预期提前了一天到货，所以在送给女儿之前，我自己玩了会。
"""
```

1.1 限制输出文本长度

我们首先尝试将文本的长度限制在30个字以内。

```
from tool import get_completion

prompt = f"""
您的任务是从电子商务网站上生成一个产品评论的简短摘要。

请对三个反引号之间的评论文本进行概括，最多30个字。

评论: ```{prod_review}```
"""

response = get_completion(prompt)
print(response)
```

熊猫公仔软可爱，女儿喜欢，但有点小。快递提前一天到货。

我们可以看到语言模型给了我们一个符合要求的结果。

注意：在上一节中我们提到了语言模型在计算和判断文本长度时依赖于分词器，而分词器在字符统计方面不具备完美精度。

1.2 设置关键角度侧重

在某些情况下，我们会针对不同的业务场景对文本的侧重会有所不同。例如，在商品评论文本中，物流部门可能更专注于运输的时效性，商家则更关注价格和商品质量，而平台则更看重整体的用户体验。

我们可以通过增强输入提示（Prompt），来强调我们对某一特定视角的重视。

1.2.1 侧重于快递服务

```
prompt = f"""
您的任务是从电子商务网站上生成一个产品评论的简短摘要。

请对三个反引号之间的评论文本进行概括，最多30个字，并且侧重在快递服务上。

评论: ```{prod_review}```
"""

response = get_completion(prompt)
print(response)
```

> 快递提前到货，公仔可爱但有点小。

通过输出结果，我们可以看到，文本以"快递提前到货"开头，体现了对于快递效率的侧重。

1.2.2 侧重于价格与质量

```
prompt = f"""
您的任务是从电子商务网站上生成一个产品评论的简短摘要。

请对三个反引号之间的评论文本进行概括，最多30个词汇，并且侧重在产品价格和质量上。

评论: ```{prod_review}```
"""

response = get_completion(prompt)
print(response)
```

> 可爱的熊猫公仔，质量好但有点小，价格稍高。快递提前到货。

通过输出的结果，我们可以看到，文本以"可爱的熊猫公仔，质量好但有点小，价格稍高"开头，体现了对于产品价格与质量的侧重。

1.3 关键信息提取

在1.2节中，虽然我们通过添加关键角度侧重的 Prompt，确实让文本摘要更侧重于某一特定方面，然而，我们可以发现，在结果中也会保留一些其他信息，比如偏重价格与质量角度的概括中仍保留了"快递提前到货"的信息。如果我们只想要提取某一角度的信息，并过滤掉其他所有信息，则可以要求 LLM 进行**文本提取（Extract）**而非概括(Summarize)。

下面让我们来一起来对文本进行提取信息吧！

```
prompt = f"""
您的任务是从电子商务网站上的产品评论中提取相关信息。

请从以下三个反引号之间的评论文本中提取产品运输相关的信息，最多30个词汇。

评论: ```{prod_review}```
"""

response = get_completion(prompt)
print(response)
```

产品运输相关的信息：快递提前一天到货。

二、同时概括多条文本

在实际的工作流中，我们往往要处理大量的评论文本，下面的示例将多条用户评价集合在一个列表中，并利用 `for` 循环和文本概括（Summarize）提示词，将评价概括至小于 20 个词以下，并按顺序打印。当然，在实际生产中，对于不同规模的评论文本，除了使用 `for` 循环以外，还可能需要考虑整合评论、分布式等方法提升运算效率。您可以搭建主控面板，来总结大量用户评论，以及方便您或他人快速浏览，还可以点击查看原评论。这样，您就能高效掌握顾客的所有想法。

```
review_1 = prod_review

# 一盏落地灯的评论
review_2 = """
我需要一盏漂亮的卧室灯，这款灯不仅具备额外的储物功能，价格也并不算太高。
收货速度非常快，仅用了两天的时间就送到了。
不过，在运输过程中，灯的拉线出了问题，幸好，公司很乐意寄送了一根全新的灯线。
新的灯线也很快就送到手了，只用了几天的时间。
装配非常容易。然而，之后我发现有一个零件丢失了，于是我联系了客服，他们迅速地给我寄来了缺失的零件！
对我来说，这是一家非常关心客户和产品的优秀公司。
"""

# 一把电动牙刷的评论
review_3 = """
我的牙科卫生员推荐了电动牙刷，所以我就买了这款。
到目前为止，电池续航表现相当不错。
初次充电后，我在第一周一直将充电器插着，为的是对电池进行条件养护。
过去的3周里，我每天早晚都使用它刷牙，但电池依然维持着原来的充电状态。
不过，牙刷头太小了。我见过比这个牙刷头还大的婴儿牙刷。
我希望牙刷头更大一些，带有不同长度的刷毛，
这样可以更好地清洁牙齿间的空隙，但这款牙刷做不到。
总的来说，如果你能以50美元左右的价格购买到这款牙刷，那是一个不错的交易。
制造商的替换刷头相当昂贵，但你可以购买价格更为合理的通用刷头。
这款牙刷让我感觉就像每天都去了一次牙医，我的牙齿感觉非常干净！
"""

# 一台搅拌机的评论
review_4 = """
在11月份期间，这个17件套装还在季节性促销中，售价约为49美元，打了五折左右。
可是由于某种原因（我们可以称之为价格上涨），到了12月的第二周，所有的价格都上涨了，
同样的套装价格涨到了70-89美元不等。而11件套装的价格也从之前的29美元上涨了约10美元。
看起来还算不错，但是如果你仔细看底座，刀片锁定的部分看起来没有前几年版本的那么漂亮。
```

```
然而，我打算非常小心地使用它
（例如，我会先在搅拌机中研磨豆类、冰块、大米等坚硬的食物，然后再将它们研磨成所需的粒度，
接着切换到打蛋器刀片以获得更细的面粉，如果我需要制作更细腻/少果肉的食物）。
在制作冰沙时，我会将要使用的水果和蔬菜切成细小块并冷冻
（如果使用菠菜，我会先轻微煮熟菠菜，然后冷冻，直到使用时准备食用。
如果要制作冰糕，我会使用一个小到中号的食物加工器），这样你就可以避免添加过多的冰块。
大约一年后，电机开始发出奇怪的声音。我打电话给客户服务，但保修期已经过期了，
所以我只好购买了另一台。值得注意的是，这类产品的整体质量在过去几年里有所下降
，所以他们在一定程度上依靠品牌认知和消费者忠诚来维持销售。在大约两天内，我收到了新的搅拌机。
"""


reviews = [review_1, review_2, review_3, review_4]

for i in range(len(reviews)):
    prompt = f"""
    你的任务是从电子商务网站上的产品评论中提取相关信息。

    请对三个反引号之间的评论文本进行概括，最多20个词汇。

    评论文本: ```{reviews[i]}```
    """
    response = get_completion(prompt)
    print(f"评论{i+1}: ", response, "\n")
```

评论1: 熊猫公仔是生日礼物，女儿喜欢，软可爱，面部表情和善。价钱有点小，快递提前一天到货。

评论2: 漂亮卧室灯，储物功能，快速送达，灯线问题，快速解决，容易装配，关心客户和产品。

评论3: 这款电动牙刷电池续航好，但牙刷头太小，价格合理，清洁效果好。

评论4: 该评论提到了一个17件套装的产品，在11月份有折扣销售，但在12月份价格上涨。评论者提到了产品的外观和使用方法，并提到了产品质量下降的问题。最后，评论者提到他们购买了另一台搅拌机。

三、英文版

1.1 单一文本概括

```
prod_review = """
Got this panda plush toy for my daughter's birthday, \
who loves it and takes it everywhere. It's soft and \
super cute, and its face has a friendly look. It's \
a bit small for what I paid though. I think there \
might be other options that are bigger for the \
same price. It arrived a day earlier than expected, \
so I got to play with it myself before I gave it \
to her.
"""

prompt = f"""
Your task is to generate a short summary of a product \
review from an ecommerce site.

Summarize the review below, delimited by triple
```

```
backticks, in at most 30 words.

Review: ```{prod_review}```
"""

response = get_completion(prompt)
print(response)
```

```
This panda plush toy is loved by the reviewer's daughter, but they feel it is a
bit small for the price.
```

1.2 设置关键角度侧重

1.2.1 侧重于快递服务

```
prompt = f"""
Your task is to generate a short summary of a product \
review from an ecommerce site to give feedback to the \
Shipping deparmtment.

Summarize the review below, delimited by triple
backticks, in at most 30 words, and focusing on any aspects \
that mention shipping and delivery of the product.

Review: ```{prod_review}```
"""

response = get_completion(prompt)
print(response)
```

```
The customer is happy with the product but suggests offering larger options for
the same price. They were pleased with the early delivery.
```

1.2.2 侧重于价格和质量

```
prompt = f"""
Your task is to generate a short summary of a product \
review from an ecommerce site to give feedback to the \
pricing deparmtment, responsible for determining the \
price of the product.

Summarize the review below, delimited by triple
backticks, in at most 30 words, and focusing on any aspects \
that are relevant to the price and perceived value.

Review: ```{prod_review}```
"""

response = get_completion(prompt)
print(response)
```

> The customer loves the panda plush toy for its softness and cuteness, but feels it is overpriced compared to other options available.

1.3 关键信息提取

```
prompt = f"""
Your task is to extract relevant information from \
a product review from an ecommerce site to give \
feedback to the Shipping department.

From the review below, delimited by triple quotes \
extract the information relevant to shipping and \
delivery. Limit to 30 words.

Review: ```{prod_review}```
"""

response = get_completion(prompt)
print(response)
```

> The shipping department should take note that the product arrived a day earlier than expected.

2.1 同时概括多条文本

```
review_1 = prod_review

# review for a standing lamp
review_2 = """
Needed a nice lamp for my bedroom, and this one \
had additional storage and not too high of a price \
point. Got it fast - arrived in 2 days. The string \
to the lamp broke during the transit and the company \
happily sent over a new one. Came within a few days \
as well. It was easy to put together. Then I had a \
missing part, so I contacted their support and they \
very quickly got me the missing piece! Seems to me \
to be a great company that cares about their customers \
and products.
"""

# review for an electric toothbrush
review_3 = """
My dental hygienist recommended an electric toothbrush, \
which is why I got this. The battery life seems to be \
pretty impressive so far. After initial charging and \
leaving the charger plugged in for the first week to \
condition the battery, I've unplugged the charger and \
been using it for twice daily brushing for the last \
3 weeks all on the same charge. But the toothbrush head \
is too small. I've seen baby toothbrushes bigger than \
this one. I wish the head was bigger with different \
length bristles to get between teeth better because \
```

```python
this one doesn't.  Overall if you can get this one \
around the $50 mark, it's a good deal. The manufactuer's \
replacements heads are pretty expensive, but you can \
get generic ones that're more reasonably priced. This \
toothbrush makes me feel like I've been to the dentist \
every day. My teeth feel sparkly clean!
"""

# review for a blender
review_4 = """
So, they still had the 17 piece system on seasonal \
sale for around $49 in the month of November, about \
half off, but for some reason (call it price gouging) \
around the second week of December the prices all went \
up to about anywhere from between $70-$89 for the same \
system. And the 11 piece system went up around $10 or \
so in price also from the earlier sale price of $29. \
So it looks okay, but if you look at the base, the part \
where the blade locks into place doesn't look as good \
as in previous editions from a few years ago, but I \
plan to be very gentle with it (example, I crush \
very hard items like beans, ice, rice, etc. in the \
blender first then pulverize them in the serving size \
I want in the blender then switch to the whipping \
blade for a finer flour, and use the cross cutting blade \
first when making smoothies, then use the flat blade \
if I need them finer/less pulpy). Special tip when making \
smoothies, finely cut and freeze the fruits and \
vegetables (if using spinach-lightly stew soften the \
spinach then freeze until ready for use-and if making \
sorbet, use a small to medium sized food processor) \
that you plan to use that way you can avoid adding so \
much ice if at all-when making your smoothie. \
After about a year, the motor was making a funny noise. \
I called customer service but the warranty expired \
already, so I had to buy another one. FYI: The overall \
quality has gone done in these types of products, so \
they are kind of counting on brand recognition and \
consumer loyalty to maintain sales. Got it in about \
two days.
"""

reviews = [review_1, review_2, review_3, review_4]
```

```python
for i in range(len(reviews)):
    prompt = f"""
    Your task is to generate a short summary of a product \
    review from an ecommerce site.

    Summarize the review below, delimited by triple \
    backticks in at most 20 words.

    Review: ```{reviews[i]}```
    """
    response = get_completion(prompt)
    print(i, response, "\n")
```

0 Soft and cute panda plush toy loved by daughter, but small for the price.
Arrived early.

1 Great lamp with storage, fast delivery, excellent customer service, and easy
assembly. Highly recommended.

2 Impressive battery life, but toothbrush head is too small. Good deal if bought
around $50.

3 The reviewer found the price increase after the sale disappointing and noticed
a decrease in quality over time.