

第5章

自然语言处理

学习目的与要求

在本章中,将探索人工智能领域中与自然语言处理相关的关键概念和技术。学习自然语言处理的目的是让计算机能够理解和处理自然语言,实现对自然语言的智能分析和应用。通过学习本章,读者将对语义分析、文本分类和情感分析、智能问答、聊天机器人、机器翻译的发展和挑战有全面的了解。这些技术和应用为文本分析、语音处理和智能交互等任务提供了有力的工具和方法。最后,本章通过介绍大语言模型,如 Transformer 模型等,进一步探索人工智能领域中的前沿技术和相关发展。

自然语言处理(Natural Language Processing, NLP)技术是与自然语言的计算机处理有关的所有技术的统称,其目的是使计算机能够理解和接受人类用自然语言输入的指令,完成从一种自然语言到另一种自然语言的翻译功能等。自然语言处理技术的研究可以丰富计算机知识处理的研究内容,推动人工智能技术的发展。

5.1 语义分析

5.1.1 语义分析的基本概念

1. 什么是语义分析

自然语言处理技术的核心为语义分析。语义分析不仅进行词法分析和句法分析这类语法水平上的分析,而且还涉及单词、词组、句子、段落所包含的意义,目的是用句子的语义结构表示语言的结构,如图 5-1 所示。语义分析可以自动完成,也可以由人完成。语义分析是对文本进行深入理解,并将其转换为计算机可以理解的形式,例如词汇表或规则集。在这个过程中需要使用大量的自然语言处理技术。

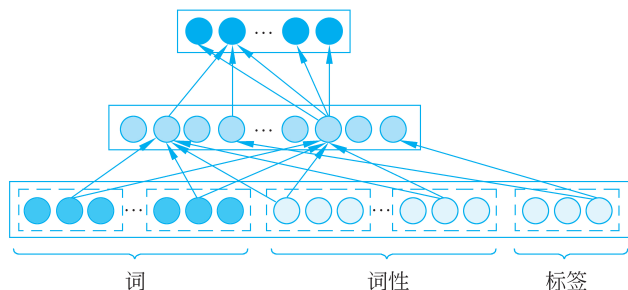


图 5-1 语义分析

2. 语义分析具体任务

1) 词法分析

词法分析包括词形分析和词汇分析两方面。一般来讲,词形分析重在对单词的前缀、后缀等进行分析,而词汇分析重在对整个词汇系统的控制,从而能够较准确地分析用户输入的自然语言的特征,最终准确地完成语义分析的过程。

2) 句法分析

句法分析是对用户输入的自然语言进行词汇、短语的分析,目的是识别句子的句法结构,以实现自动句法分析的过程。

3) 语用分析

语用分析相对于语义分析又增加了对上下文、语言背景、语境等的分析,即从文章的结构中提取出意象、情感等附加信息,是一种更高级的语言学分析。它将语句中的内容与现实生活中的细节关联在一起,从而形成动态的表意结构。

4) 语境分析

语境分析主要是指对文章之外的大量“言外之意”进行分析,以便更准确地解释语言的技术。

5) 自然语言生成

人工智能驱动的引擎能够根据收集的数据生成描述,将数据转换为语言,在人与技术之间创建无缝交互。

自然语言生成软件接受结构化表示的语义,以输出符合语法的、流畅的、与输入语义一致的自然语言文本。早期大多采用管道模型研究自然语言生成,管道模型根据不同的阶段将研究过程分解为如下 3 个子任务:

- 内容选择。决定要表达哪些内容。
- 句子规划。决定篇章及句子的结构,进行句子的融合、指代表述等。
- 表层实现。决定选择什么样的词汇实现一个句子的表达。

早期基于规则的自然语言生成技术在每个子任务上均采用不同的语言学规则或领域知识,以实现从输入语义到输出文本的转换。

基于规则的自然语言生成系统存在很多不足之处。近几年来,学者们开始了基于数据驱动的自然语言生成技术研究,从浅层的统计机器学习模型到深层的神经网络模型,对自然语言生成过程中每个子任务的建模以及多个子任务的联合建模开展了相关的研究,目前主流的自然语言生成技术是基于数据驱动的自然语言生成技术和基于深度神经网络的自然语言生成技术。

3. 语义分析与其他自然语言处理技术的区别

相比于其他自然语言处理技术(如词法分析和句法分析),语义分析更加复杂和具有挑战性。词法分析和句法分析主要关注明确单词和句子的结构;而语义分析则需要更加深入地理解这些结构,并确定其背后的含义。

通过语义分析,可以使计算机系统具有更好地理解 and 推断人类语言的能力,从而实现各种应用场景,如智能问答、机器翻译和智能客服等。

4. 中文分词

中文分词是计算机根据语义模型自动将汉字序列切分为符合人类语言理解模式的词汇。分词就是将连续的文字序列按照一定的规范切成词的过程。

在英文中,单词之间是以空格作为自然分隔符的;而中文只能够通过字、句和段落进行简单的划分,中文词没有形式上的分隔符。虽然英文也存在短语划分的问题,不过在词这一层面上,中文比英文要复杂得多、困难得多。

5.1.2 语义分析的主要应用场景

本节对语义分析的主要应用场景进行介绍。

1. 情感分析

情感分析是一种基于自然语言处理的技术,可以对文本进行情感分类和情感极性分析。随着大数据和机器学习的快速发展,情感分析的应用已经扩展到了商业营销、舆情监测、客户服务等各个领域。在商业营销中,情感分析可以帮助企业了解目标受众群体,提高营销效果,增强品牌声誉。

在社交媒体上,人们发布的内容越来越多,这为企业提供了一个机会,可以通过情感分析了解公众关注的话题,从而制定相应的营销策略。例如,当某个品牌推出新产品时,情感分析可以自动识别用户的评论和反馈,从而了解用户对产品的评价和态度。如果大多数用户反馈积极,企业可以进一步加大宣传力度,提高产品的知名度和销售额;如果用户反馈较为负面,企业可以及时改进产品,避免口碑受损。

除了营销领域,情感分析还可以应用于舆情监测和客户服务。在舆情监测方面,情感分析可以自动识别新闻报道、社交媒体评论等文本中的情感色彩,了解公众对某个事件或话题的态度和情感倾向。这对企业制定公关策略和舆情应对具有重要意义。在客户服务方面,情感分析可以自动识别客户的反馈和抱怨,及时解决客户问题,提高客户满意度和忠诚度。

尽管情感分析在商业营销中应用广泛,但是它也存在一些限制和挑战。首先,情感分析算法的准确性受到文本质量和语言表达的影响。例如,一些文本可能存在歧义或语义混淆,从而影响情感分析的准确性。其次,情感分析也面临情感分类的主观性和不确定性的挑战。同一段文本可能会被不同的人或算法分类为不同的情感极性,这需要人工干预和进一步优化算法。

总的来说,情感分析在商业营销中的应用前景广阔,但是需要不断优化算法和提高准确性,以更好地服务于企业和公众。

2. 垃圾邮件过滤

随着互联网的普及,人们经常用电子邮件进行沟通和工作。但是,随之而来的问题是垃圾邮件的数量不断增加,给人们带来了很大麻烦。垃圾邮件不仅会浪费人们的时间和精

力,还可能包含恶意代码和网络钓鱼链接,从而导致安全风险。因此,通过语义分析技术过滤垃圾邮件是非常必要的。

语义分析是一种自然语言处理技术,可以自动理解和分析文本中的意义,并从中提取有用的信息。在电子邮件过滤中,语义分析可以分析每个电子邮件的主题、内容、发送者和附件等信息,以确定它们是否是垃圾邮件。根据分析结果,系统可以将垃圾邮件自动移到垃圾箱中,同时保留正常的邮件。

为了实现高效的语义分析,机器学习算法是必不可少的。机器学习算法可以自动学习并识别垃圾邮件中的特定单词、短语或特征,并将它们与正常邮件进行区分。为了提高准确性,系统需要不断地从新的正常邮件和垃圾邮件中学习,并根据学习结果进行调整和改进。

然而,由于黑客不断改进他们的攻击方式,传统的语义分析技术可能无法完全识别新型的垃圾邮件。因此,开发新的应对策略也是必要的。一种常见的策略是将多种技术相结合,例如黑名单、白名单、规则过滤和机器学习等,以提高垃圾邮件的过滤能力。同时,用户也应该保持警惕,避免点击不明来源的链接或下载可疑附件,以确保网络安全。

语义分析技术为用户提供了一种过滤垃圾邮件的高效方法。系统通过不断学习和改进可以更加准确地识别垃圾邮件,提高工作效率,保障网络安全。

3. 新闻分类

在信息时代,新闻报道是人们获取信息的重要途径之一。然而,传统的新闻报道方式需要耗费大量的人力、物力和时间,而且容易出现信息滞后和质量不稳定等问题。随着互联网的发展,越来越多的新闻信息以数字化的形式出现,这也为基于语义分析技术的新闻分类系统提供了契机。

语义分析技术是一种自然语言处理技术,可以自动理解和分析文本中的意义,并从中提取有用的信息。在新闻分类系统中,语义分析可以分析新闻报道的标题、内容、发布时间、来源等信息,并根据这些信息将新闻分到相应的类别中。例如,一篇有关某个公司业绩的新闻报道可以被自动分类到“财经”类别中,一篇有关国际政治的新闻报道可以被自动分到“国际”类别中,这样可以大幅提高新闻报道的效率和质量。

使用语义分析技术进行新闻分类可以带来多方面的好处。首先,它可以大大减少人工编辑的工作量,提高新闻报道的效率和速度。其次,它可以让用户更快捷地找到自己感兴趣的新闻内容,提高用户体验。此外,新闻分类系统还可以帮助新闻机构更好地了解用户的兴趣和需求,为后续的新闻报道提供参考和指导。

语义分析技术在新闻分类中的应用不局限于单一领域,还可以应用于多领域的信息分类。例如,一家在线新闻门户网站可以利用语义分析算法将所有涉及某个公司的新闻内容组合在一起,或者根据某个主题将所有相关报道整理成一个专题。这样可以让用户更加方便地获取相关信息,也可以帮助新闻机构更好地了解用户的兴趣和需求。

除了新闻分类,语义分析技术还可以应用于新闻推荐和舆情分析等领域。例如,在新闻推荐方面,语义分析技术可以根据用户的历史浏览记录和关注点,为用户推荐最相关的新闻内容。在舆情分析方面,语义分析技术可以根据用户在社交媒体上的言论和情感倾向,分析公众舆论和情感趋势。

然而,语义分析技术在新闻分类中的应用也存在一些挑战和难点。首先,语义分析技

术需要大量的数据进行训练和优化,这需要投入大量的时间和资源。其次,语义分析技术可能存在误判或漏判的问题,尤其是在新闻报道中存在大量的歧义和复杂性的情况下。因此,为了提高语义分析的准确性和可靠性,需要不断地进行技术创新和优化。

基于语义分析技术的新闻分类系统已经成为新闻报道的新趋势。通过语义分析技术的应用,可以大大提高新闻报道的效率和质量,让用户更加方便地获取相关信息,同时也可以帮助新闻机构更好地了解用户的兴趣和需求。虽然在应用中面临一些挑战和难点,但随着技术的不断发展和创新,相信语义分析技术在新闻分类中的应用会越来越普及和成熟。同时,除了新闻分类,语义分析技术还有着广泛的应用前景,将在更多的领域为人们带来便利和好处。

4. 智能客服

随着市场竞争日趋激烈,企业越来越需要拥有一种更为有效的方式处理客户服务请求。传统的人工客服可能无法满足快速增长的客户需求,因此利用语义识别技术提供智能客服成为一种可行的解决方案,不仅可以提高客户满意度,还能有效降低企业的运营成本。

智能客服的核心是利用语义分析技术,通过自然语言处理技术和人工智能技术的应用,对客户的问题进行准确的理解和回答。相比于传统的常见问题解答,智能客服通过将输入的问题与现有的知识库进行比对,可以提供更加准确和个性化的答案,从而帮助客户解决问题。无论是简单的常见问题还是复杂的技术支持,智能客服都能够快速作出反应并提供专业的解决方案。

与此同时,智能客服还可以通过语义分析技术识别客户的语气、情绪和问题背后的真正需求。这使得智能客服更加接近人类的性质,能够更好地理解客户的意图和情绪,从而提供更加贴近客户期望的服务。无论客户是抱怨、咨询还是寻求帮助,智能客服都能够敏锐地捕捉到客户的情绪变化,并采取相应的措施,以提供个性化和温暖的服务体验。

智能客服的优势还体现在其可以实现不间断的在线服务。智能客服可以随时在线,始终准备着为客户提供帮助。无论是在白天还是深夜,无论是在工作日还是节假日,智能客服都能够快速响应客户并提供解决方案。这种不间断的在线服务不仅提高了客户的满意度,也增强了企业的竞争力。

当然,智能客服并不是要取代人工客服,而是将其作为补充和升级。在一些复杂问题或涉及人的情绪的情况下,人工客服的作用仍然是不可替代的。智能客服可以通过语义分析技术将问题分类和转接给相应的人工客服,以确保问题得到妥善解决。

随着市场竞争的加剧,智能客服以其准确、个性化、全天候的优势成为企业处理客户服务请求的重要工具。通过语义识别技术,智能客服能够理解客户的问题和需求,提供快速而专业的解决方案,并在情感和语气上与客户产生共鸣。智能客服的出现为企业提供了一种更加高效和智能的客户服务方式,必将在市场竞争中占据重要的位置。

5.2 文本分类和情感分析

5.2.1 文本分类和情感分析的基本概念

1. 背景介绍

文本分类是自然语言处理最基础、最核心的任务,或者说,几乎所有自然语言处理任务

都是分类任务,或者涉及分类的概念。例如,分词、词性标注、命名实体识别等序列标注任务其实就是 Token 粒度的分类;再如,文本生成其实也可以理解为 Token 粒度在整个词表上的分类任务。

情感分析是根据输入的文本、语音或视频,自动识别其中的观点倾向、态度、情绪、评价等,广泛应用于消费决策、舆情分析、个性化推荐等商业领域。情感分析包括篇章级、句子级和对象或属性级。

(1) 篇章级。整个文本作为输入,输出整体的情感倾向。例如,给定一段评论,判断整体是正向(积极、褒义等)还是负向(消极、贬义等)。

(2) 句子级。输入的是一个句子,其他的同上。

(3) 对象或属性级。

判断给定文本中某对象或属性的情感倾向。若给定文本、对象或属性,判断对象或属性的情感倾向;若只给定文本,需要首先提取其中的对象或属性,然后再判断情感倾向。

例如,某家餐饮店下的评论:“服务非常赞,但味道很一般。”这句话其实表达了两个意思,或者说两个对象(属性)的评价,需要输出类似<服务,正向>和<口味,负向>这样的结果(<属性,情感倾向>二元组),或者再细一点,加入用户的观点信息(<属性,观点,情感倾向>三元组),输出<服务,赞,正向>和<口味,一般,负向>。这样的结果才更有实际意义。

从自然语言处理的角度看,情感分析包括两个任务:文本分类和实体对象或属性抽取,而这也正好涵盖了自然语言处理的两类基本任务——序列分类(sequence classification)和 Token 分类(Token classification)。两者的区别是:前者输出一个标签,后者每个 Token 输出一个标签。

2. 基本流程

一般来说,自然语言处理任务的基本流程主要包括 5 个步骤,如图 5-2 所示。

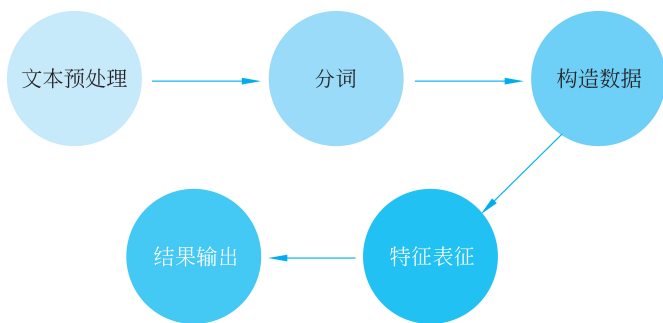


图 5-2 自然语言处理任务的基本流程

1) 文本预处理

文本预处理主要是根据任务需要对输入文本进行的一系列处理,主要包括以下几个操作:

(1) 文本清理。去除文本中不参与处理的字符,例如网址、图片地址,无效的字符、空白、乱码等。

(2) 标准化。主要是将不同的形式统一化,例如英文大小写、数字形式、英文缩写、日期格式、时间格式、计量单位、标点符号等。

(3) 纠错。识别并纠正文本中的错误,包括拼写错误、词法错误、句法错误、语义错

误等。

(4) 改写。包括转换和扩展。转换是将输入的文本或查询转换为同等语义的另一种形式,例如将拼音(或简拼)转换为对应的中文。扩展主要是将和输入文本相关的内容一并作为输入,常用在搜索领域。

需要注意的是,这个处理过程并不一定是按照上面的顺序从头到尾执行的,可以根据需要灵活调整,例如先纠错再标准化或将标准化放到改写里面。

2) 分词

分词主要目的是将输入的文本转换为 Token,它涉及后续将文本转换为向量。一般主要从 3 个粒度对文本进行切分:

(1) 字级别。英文就是字母级别,操作起来比较简单。

(2) 词级别。英文不需要分词。中文分词主要是分割语义,降低不确定性。

(3) 子词。在英文中很常见,当然中文也可以切分子词。子词是介于字级别和词级别中间的粒度。切分子词的主要目的是将一些统一高频的形式单独切分出来,例如英文中的前缀 de 或者表示最高级的后缀 est 等。子词一般是在大规模语料上通过统计频率自动学习到的。

3) 构造数据

文本经过分词后会变成一个个 Token,接下来就是根据后续需要将 Token 转换为一定形式的输入,可能就是 Token 序列本身,也可能将 Token 转换为数字,或者再加入新的信息(如特殊信息、Token 类型、位置等)。这里主要考虑后两种情况。

(1) Token 转换为数字。将每个 Token 转换为一个整数,一般就是它在词表中的位置。根据后续模型的不同,可能会有一些特殊的 Token 加入,主要用于分割输入,其实是一个标记。有两个常用的特殊 Token 需要稍加说明: <UNK> 和 <PAD>,前者表示未识别的 Token,后者表示填充用 Token。在实际应用中往往会批量输入文本,而文本长度一般是不相等的,这就需要将它们变成统一的长度,也就是把短的用 <PAD> 填充。

(2) 加入新的信息。又可以进一步分为在文本序列上加入新的信息和加入和文本序列平行的信息。

4) 特征表征

文本特征是指将文本转换为可以用于机器学习算法的向量表示。常见的文本特征表示方法有 One Hot 编码、TF-IDF 和 Embedding。

(1) One Hot 编码。

One Hot 编码是指将每个单词或字符都转换为一个独立的二元特征,即,如果该单词或字符出现在文本中,则对应的特征值为 1;否则为 0。例如,假设有一个包含 3 个句子的文本集合。

句子 1: I love machine learning

句子 2: I hate math

句子 3: Machine learning is fun

那么就可以将这 3 个句子转换为 One Hot 编码的特征向量表示,如图 5-3 所示。

(2) TF-IDF。

TF-IDF 是一种统计方法,用于评估一个词对于一个文档集合中的一个文档的重要程

单词/字符	I	love	machine	learning	hate	math	is	fun
句子1	1	1	1	1	0	0	0	0
句子2	1	0	0	0	1	1	0	0
句子3	0	0	1	1	0	0	1	1

图 5-3 One Hot 编码示例

度。TF-IDF 由两部分组成：词频(Term Frequency, TF)和逆文档频率(Inverse Document Frequency, IDF)。词频指某个词在该篇文档中出现的次数。逆文档频率是为了减少常见词对文档的影响而使用的一种方法,具体来说,逆文档频率的值等于总文档数除以包含该词语的文档数再取对数。

TF-IDF 算法的输出结果是一个矩阵,其中每一行对应一个文档,每一列对应一个词或字符,矩阵中的一个元素表示该词或字符在该文档中的 TF-IDF 值。

(3) Embedding。

Embedding 是一种将词或字符映射到高维空间的技术,使得词或字符的距离可以反映它们之间的语义关系。常见的 Embedding 模型有 Word2Vec、GloVe 和 FastText 等。这些模型都是基于神经网络的模型,通过学习词或字符在上下文中的分布构建词或字符的向量表示。

Embedding 的输出结果是一个矩阵,其中每一行对应一个文档,每一列对应一个 Embedding 向量的维度,矩阵中的每个元素表示该文档中对应词或字符的 Embedding 向量。

5) 结果输出

当用户的输入是一句话(或一段文档)时,往往需要得到整体的向量表示。在 Embedding 方法出现之前虽然也可以通过频率统计得到,但难以进行后续的计算;在 Embedding 出现之后,方法就有很多了,其中最简单的是对每个词的向量求和再求平均值。

在得到整个句子(或文档)的向量表示后,如何得到最终的分类呢?很简单,通过矩阵乘法将向量转换为一个类别维度大小的向量。以二分类为例,首先将一个固定维度的句子或文档向量变为一个二维向量,然后将该二维向量通过一个非线性函数映射成概率分布。

3. 常用算法

文本分类和情感分析是自然语言处理中两个常见的任务。下面介绍它们的常用算法。

1) 文本分类常用算法

文本分类是指将给定的文本划分到预定义类别中,例如,将邮件分为垃圾邮件或非垃圾邮件,将新闻文章分类为政治、体育或娱乐等类别。

常用的文本分类算法有以下 3 个:

(1) 朴素贝叶斯分类器。基于贝叶斯定理,计算每个类别的概率,并选择具有最高概率的类别作为分类结果。

(2) 支持向量机分类器。通过构建一个超平面将数据集分成不同的类别。

(3) 决策树分类器。通过对特征进行递归二分,生成一棵决策树,从而实现分类。

2) 情感分析常用算法

情感分析是指对文本进行情感极性的分类,通常分为正面、负面和中性 3 类。

常用的情感分析算法有以下 3 种:

- (1) 基于规则的方法。通过人工制定一些规则判断文本的情感极性。
- (2) 基于情感词典的方法。构建一个包含情感词汇和对应情感极性的词典,根据文本中出现的情感词汇及其情感极性计算文本的情感极性。
- (3) 基于机器学习的方法。通过训练一个分类器判断文本的情感极性。常用的分类器有朴素贝叶斯分类器、支持向量机分类器和神经网络等。

5.2.2 文本分类和情感分析的应用场景与展望

1. 实际应用

这里主要探讨在实际应用过程中对方法和模型的取舍和选择。

(1) 规则与模型。纯规则、纯模型和两者结合的方法都有。规则可控,但维护不容易,尤其当规模变大时,规则重复、规则冲突等问题就会出现;模型维护简单,但可能需要增加语料,另外过程也不能干预。在实际应用中,简单任务可以使用纯模型;复杂可控任务可以使用模型+规则组合,模型负责通用大多数,规则覆盖长尾个案。

(2) 深度学习与传统机器学习。当业务需要可解释时,可以选择传统的机器学习模型;当没有这个限制时,应优先考虑深度学习模型。

(3) 简单与复杂。当任务简单、数据量不大时,可以用简单模型(如 TextCNN),此时用复杂模型未必效果更好;但是当任务复杂、数据量比较大时,复杂模型的性能远远超过简单模型。

总而言之,使用什么方案要综合考虑具体的任务、数据、业务需要、产品规划、资源等多种因素后确定。

2. 情感分析的未来

情感分析的未来一定是围绕着更深刻的语义理解发展的。从情感分析目前的发展看,主要有以下 3 个方向。

1) 多模态

多模态是指多种不同形态的输入结合,包括文本、图像、声音、视频等。这也是目前比较前沿的研究方向。

2) 深度语义

深度语义是指综合考虑多种影响因素,包括环境、上下文、背景知识。这和多模态类似,也是尽量将场景真实化。

环境指的是对话双方当前所处的环境。例如,现在是冬天,但用户说“房间怎么这么热?”其实可能是因为房间空调或暖气开太高。如果不考虑环境,可能就会难以理解这句话。

上下文指的是对话中的历史信息。例如,开始对话时用户说“今天有点感冒”,后面如果说“感觉有点冷”,那可能是生病导致的。如果没有这样的上下文,对话可能看起来就不合理,对情感的理解和判断也会不准确。

背景知识是指关于世界认知的知识。

3) 多方法

多方法是指综合使用多种方法,包括知识图谱、强化学习和深度学习。知识图谱主要是对世界万物及其基本关系进行建模,强化学习则是对事物运行的规则进行建模,而深度学习主要考虑实例的表征,多种方法组合成一个立体、完整的系统。

3. 应用场景

文本分类和情感分析是自然语言处理中两个常见的任务，它们在很多领域都有广泛的应用。下面简要介绍它们的应用场景。

1) 文本分类

文本分类可以应用于以下场景：

- (1) 垃圾邮件过滤。例如，将收到的邮件分为垃圾邮件和非垃圾邮件。
- (2) 新闻分类。例如，将新闻文章分为政治、体育、娱乐等类别。
- (3) 产品分类。例如，将电商平台上的商品分为服装、数码、家居等类别。
- (4) 问题分类。例如，将用户反馈的问题分为技术支持、账户问题、产品建议等类别。

2) 情感分析

情感分析可以应用于以下场景：

- (1) 舆情监测。对社交媒体上的评论等进行情感分析，了解公众对某个事件或话题的态度。
- (2) 用户评论分析。对用户在该电商平台上的评论进行情感分析，了解用户对产品或服务的评价。
- (3) 基于情感的推荐。根据用户历史评价数据，对其进行情感分析，推荐符合其喜好的产品或服务。
- (4) 品牌管理。对品牌在社交媒体上的声誉进行情感分析，了解公众对品牌的态度，及时采取应对措施。

5.3 智能问答

5.3.1 智能问答系统的基本概念

1. 智能问答系统是什么

智能问答系统是自然语言处理领域中一个很经典的问题，它可以用来回答人们以自然语言形式提出的问题。这需要对自然语言问题进行语义分析，包括关系识别、实体连接、形成逻辑表达式，然后到知识库中查找可能的备选答案，再通过排序机制给出最佳答案。

智能问答系统通常分为任务型系统、解决型系统、闲聊系统。这 3 种类型的机器人分别应用在不同的场景：

- (1) 任务型系统主要用于完成用户的某些特定任务，例如天气咨询、买机票、充电费等。
- (2) 解决型系统主要用于解决用户的问题，例如商品购买咨询、商品退货咨询等。
- (3) 闲聊系统主要用于深入地和用户进行无目的交流。

2. 智能问答系统是如何工作的

智能问答系统是一种基于人工智能和自然语言处理技术的软件系统，它可以理解用户提出的问题并给出相应的答案。下面是智能问答系统的工作流程。

1) 语言理解

当用户输入一个问题时，智能问答系统会先进行语言理解，将自然语言转换为计算机