

大数据下人工智能技术在信息检索中的应用探析

文 / 杨凡

随着社会经济的发展，人工智能技术得到了新的发展，在各行业中得到了广泛应用，并受到使用者的一致好评。人工智能技术在社会中的应用为人们生产生活提供了便利。信息检索是人们获取信息的主要方法和手段，将人工智能技术应用到信息检索中，可以为使用者提供全新的使用体验。不仅可以节约使用者的信息检索时间，使其在短时间内获得所需信息，同时还可以增强人工智能技术在信息检索中的应用效果。本文就大数据时代背景下对人工智能技术在信息检索中的应用策略进行探究，旨在减少使用者的信息使用障碍，进而实现社会的可持续发展。

人工智能技术在大数据时代背景下得到了良好的发展。人工智能是建立在多个学科之上的技术，主要包括互联网技术、数据分析以及网络通信等，范围甚广。经分析，人工智能技术的应用可以提高人们生产效率，在生产生活各个领域，运用人性化思维操作机械设备以及控制系统运行。信息技术的发展带动了人工智能技术的创新和发展，通过虚拟系统、语音操作等使各行业发生了巨大的变化。本文就大数据背景下，对人工智能在信息检索方面做出的贡献进行分析和探究，为行业从业者提供发展思路。

大数据背景下人工智能信息检索技术

大数据

随着信息技术的发展，时代背景下信息资源规模扩大，信息种类逐渐增加。人们面对数量庞大的信息资源无法及时整合。因此，信息行业从业者应及时意识到技术创新的重要性，结合时代发展需求，积极进行技术创新，以便于对丰富多样的数据信息进行归纳和梳理。大数据主要有以下几个特点：首先，大数据信息来源丰富且种类较多，并且随着时代的发展，图文类信息数量逐渐增加；第二，大数据具有较强的时效性，

信息更新速度较快；第三，大数据数量庞大，存储起来存在一定难度；最后，由于大数据更新速度较快，对存储软件以及存储方式提出了新的要求，传统存储方式已经不能满足数据分析的要求。

人工智能

人工智能技术作为科学发展中应用较为广泛的一种技术，走在时代前沿。主要思想是运用人的思维对数据信息进行分析和处理。将其应用到社会各个领域，均可以取得良好的效果。这门学科是一项综合性学科，其中涉及数学、心理学以及计算机等多门学科的内容。随着人工智能技术的应用，研究深度逐渐增加，包括算法分析、语音识别以及网络分析等。该技术在信息检索领域主要是记忆检索、智能分析以及体系建立等信息检索办法。

信息检索

信息检索是将信息系统中的数据、信息按照一定管理模式进行整合。在信息检索过程中，使用者通过输入关键词，在信息库中筛选中自己需要的信息，进而完成检索任务。信息检索主要有两方面内容，第一是通过检索关键词对信息库中的数据信息进行搜索；第二则是将信息按照一定顺序进行

整合，对数据信息进行分类，随后进行信息查询。

信息检索面临的问题

当前主要使用的信息检索主要有两种方式：第一个是将网站按照一定功能和作用进行归类分析，对每一网站功能进行简单描述；第二是运用全文检索技术，即对全部文本进行处理，通过对文本内容的掌握，设计文档索引，提高信息检索效率。

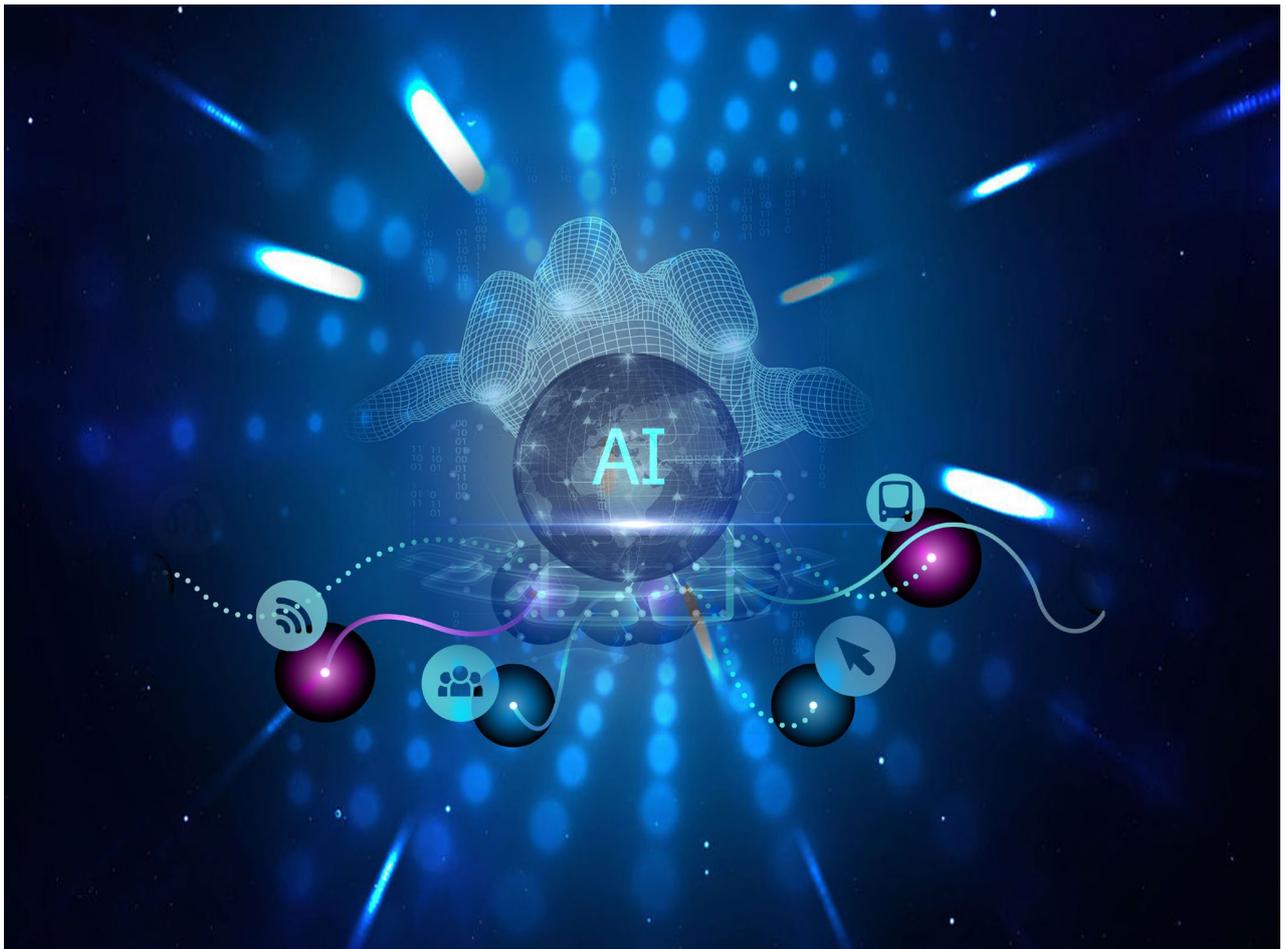
以往网站分类过程中，主要是通过人力进行网站功能的总结和梳理。该技术为信息索引提供了便利，受到人们的欢迎。但是由于相关人员在网站描述过程中不够细致，未能对网站的内容功能进行掌握。这就导致使用者在使用过程中无法通过简单的关键词查询内部信息，可能会出现信息缺失和遗漏的情况。

全文索引主要是通过网页进行信息查询。一般来讲，只

要网页上出现了关键词，那么运用全文索引就可以将网页查询出来。但是这种查询方式有两个问题，首先，用户信息索引过程中，无法通过关键词的简单串联表达出真正想要表达的内容，检索存在障碍；另外一个则是语言差异问题，人类的语言习惯和语言表达形式，随着时间、地域的改变有所差异。同一句话可以用不同的语序和词汇来表达。互联网的发展为人工智能技术提供了一个广阔的舞台，为了提高信息检索效率，有效解决实际问题，需要将人工智能技术应用到其中。

人工智能技术在信息检索中的体现

智能检索指的是依据用户在网络上提供的信息，进行信息收集分析和分类处理，在此过程中，完成对用户身份以及偏好的记录和筛选。如果用户提供的信息不完整，则需要依据用户提供的现有信息进行分析和判断，确定用户的需求，实现信息存储和筛选。



信息过滤技术

信息过滤技术是指对网络平台中的内容进行过滤和筛选。主要应用在两方面。第一，在网络平台运行过程中，其中蕴含大量的数据信息。在这些信息中，好坏参半，通过信息过滤，能够将信息按照质量进行排序。使用者在进行信息检索时，优质信息会出现在检索页面的前几列，保证信息的准确和科学性。第二则是网络内容进行管理。通过杀毒软件对垃圾信息进行拦截，净化网络环境。就以往使用的网络过滤方法而言，虽然使用起来较为简单，但是模式单一，不能满足大数据背景下对网络信息过滤的要求。往往是在信息库内进行信息筛选，无法对文本进行具体分析。而人工智能技术下的过滤技术能够弥补传统方法的不足，迅速对文档内容进行智能识别和筛选，为使用者提供高效、优质的信息检索服务。

自然语言处理技术

语言处理是人工智能领域中的重要研究内容。通过将人类的语言习惯录入到网络系统中，对人们的语言组成以及语言表达进行分析，掌握人类的语言规律后，将其运用到计算机语言体系中，通过网络代码的编制实现人类与计算机的有效沟通。另外，将复杂的语言融入到信息检索体系中，帮助网络平台对使用者输入的词汇以及语言进行识别，判断其是否为不良信息。如果为不良信息，则对其进行过滤，如果符合网络语言使用标准，则对文本内容进行分析和处理，找到使用者的目标网页和文本内容。



语音识别技术

语音识别技术主要是指运用一定科学技术将语音信号录入到信息系统中，涉及发声、听觉以及概率论等学科内容，应用较为广泛。应用该技术，可以实现网络平台与人类的实时对话，网络系统能够将人们输出的语言进行识别，及时过滤掉不健康的信息。同时提取有效语音信息，将其作为关键词进行信息检索。其工作原理是将语音中的信息进行分段、分类处理，将语音进行整体分析后，重新排列组合，使其与语音系统相匹配，能够第一时间对使用者输出的语音内容进行应用和管理。在语音识别过程中，由于当前信息技术未能对不同国家、不同地区的语言进行收录，这就导致在语音识别中，系统只能对标准语言进行判断和应用，无法在第一时间对方言以及特色语言进行识别和检索。

图像识别与视频检索技术

图像和视频检索又被称为机器视觉，主要是将图像信息第一时间转化为图像信号，通过对信号的识别，完成信息检索。每一幅图像均有其与众不同的特点，在图像识别前，首先需要人工将图像的显著特征进行系统录入，完善信息图像库。随后，图像信息检索过程中，通过对图像的整体识别，掌握图像的显著特点，在有限时间内获取最多信息，进而完成对事物的分辨。在图像录入过程中，是依据人眼的观察规律进行的，人们对于事物的观察具有一定特点，而信息检索则是根据人的观察习惯记录图像信息。为了提高信息检索效率。系统在识别前，需要对海量的数据信息进行搜集，运用一定计算原理和数学方法进行学习。在信息录入前，需要对图像信息进行分类，依据分类标准存储信息，进而实现信息检索。

视频检索技术与图像检索技术存在一定差异性。主要表现在对目标的观察、记录，识别和分析。由于视频中内容大多为移动物体，通过对其运动特点以及色彩构成等参数进行跟踪分析，组成特征向量。对视频播放过程进行图像截取，进而完成对视频内容的分类和检索。

智能技术在信息检索的应用策略

基于本体论

在以往信息检索过程中，使用最为频繁的是关键词检索和分类检索，这两种方法虽然操作起来较为简单，但是使用效果不尽如人意。这是因为这些方法脱离实际，各自为战，缺乏联系，人们对于这些概念的理解仍然停留在表面。本体论是指将相关内容进行联系，形成完整的检索体系。这种体系的特点在于用一种属性来形容相似信息的特点，可以从语义网络的角度进行表示，网络节点代表着一个概念，而概念间的连线代表着概念间的联系。通过树状图的形式，来体现

数据库内容的关系。在具体检索过程中,主要有两种检索方式,第一是盲目搜索,即将自己想要获取的知识输入到信息搜索条中,进行全面搜索;这种搜索方式对于词汇敏感性不强,无法进行智能分析;而启发式搜索则与之有所不同,在搜索过程中,将搜索内容的关键点和特点进行输入,按照自己的理解不断调整搜索方向和方式,进而寻找到最优搜索办法。这种搜索方法不仅更加人性化,体现了搜索的智能化,同时还可以缩短使用者的搜索时间,提高其信息检索效率。

基于神经网络

神经网络是由一个个神经元组合而成的,体现了生物体的思考过程和思考形式。随着人工智能技术的深入发展,人工神经网络应运而生。这种网络图能够提高信息检索效率,同时能够缓解检索平台的工作压力。人工神经网络图并非一成不变,而是通过神经元位置的调整而改变的。除此之外,人工神经网络主要是对人体的思考形式进行模仿,对人类思维进行感知,这种管理模式能够同时处理大容量的数据信息,联想能力和学习能力较强。当前运用的人工神经网络主要有两种,分别是无反馈网络和反馈网络,与无反馈网络相比,反馈网络能够按照使用者的检索记录和检索方式,对用户使用情况进行实时跟踪,将整理得到的内容反馈到上一管理体系,为使用者提供更加准确的搜索结果。反馈检索有正例和反例两种。使用者信息检索过程中,反馈网络开始运行,对于使用者输入的词条以及检索内容,平台进行整合和分析,根据一定函数图像,选取出一定范围,进而在范围内选择检索结果,为用户呈现,用户不满意可以刷新检索页面,系统内部进行反馈,为用户提供多元化检索结果,直至用户得到满意的检索结果为止。

优化检索质量的方法

存在的问题

在信息收集前,使用者由于生活环境以及职业年龄等特征不同,对于信息检索需求也不一样。行业专业可以解决用户的这些问题。但是计算机系统与人工不同,虽然运用了人工智能技术,毕竟不是真正的人,处理问题的思维与人类还是存在差异。在用户定位时,存在一定难度。不同领域的专家对各自领域均有不同的见解和想法,但是计算机内未能及时将专家的专业技能进行掌握,这也就导致无法对专家信息检索库进行建立。当前主要存在的检索问题是分类技术不够精细,无法通过简单的检索条对检索对象实现全面了解,一旦信息数量大幅度增加,检索准确率会下降,增加使用者的检索难度。

解决办法

智能代理技术是人工智能技术的重要研究方向。一般来讲,人类也在智能代理的范畴内,包括常见的智能机器人和移动机器人。而从另一个角度来讲,智能代理主要指的是软件信息机器人,代表着用户和网络程序。总之,智能代理在工作时,不需要人类的帮助和干预就可以完成预期任务。智能代理一般有四个特征,分别是代理性、智能性、机动性和个性化。

代理性主要是对智能代理的自主工作能力以及协调能力进行描述,即可以离开人们的干预进行自动化工作。但是代理在系统中,必须控制自己的行为。如果其他代理想要干预其操作,需要对该代理提出请求,代理依据自己的能力判断是否接受任务。智能性是指代理的学习能力,代表了用户完成任务的能力,例如帮助用户用信息语言表达对信息处理和数据计算的需求,帮助用户减少语言障碍,能够提高检索效率。与此同时,通过对用户的检索过程进行分析,判断用户的兴趣和喜好,推测用户的使用用途。机动性也叫做智能代理的迁移能力。从一台机器上对用户的使用习惯进行记录,将运行和操作系统转移到另一台机器上。机器间互相连通能够对用户的使用差异进行比较,进行深度分析。因此,必要时,智能代理需要与其他代理进行交流和沟通,提高信息检索的有效性。最后,智能代理具有个性化特点,在这种使用模式下,用户在浏览过程中可以优化检索结果,使其信息更加清晰。

以往的信息检索引擎不够主动,只有在输入搜索词条时才会启动和运行。而智能代理技术更加个性化和人性化,能够为使用者量身定做搜索系统。对于用户的信息搜索情况进行实时监测和记录分析,为用户提供准确的搜索结果,满足用户需求。

人工智能技术的发展对信息检索产生了巨大的影响。将信息技术与智能技术相结合,能够提高信息检索效率和质量。在大数据时代下,对于庞大的信息数据进行科学整理和分类,进而实现人性化、智能化管理。总之,人工智能技术与信息检索技术相辅相成,互相促进,有利于加快互联网信息化建设,为人们提供优质信息化服务。(本文为作者为四川九洲视讯科技有限责任公司高级工程师)科