

智能化色谱分析软件的研究与应用探讨

王亚敏 杨海鹰

(中国石油化工股份有限公司石油化工科学研究院 北京 100083)

E-mail: wangym2@ripp-sinopec.com

摘要 介绍了智能化色谱软件的概念,比较了传统分析软件与智能软件在开发流程上的区别,给出了对于一些常规分析方法由传统数据处理软件向智能化软件发展的必要性的原因。结合气相色谱在汽油单体烃分析、模拟蒸馏及炼厂气组成三大领域的实际应用案例,探讨了此类软件的开发思路及应用体会。

关键词 智能化;软件;气相色谱

中图分类号 O657.7⁺1

Discussing and Application for Intelligent Chromatogram Software

Wang Yamin, Yang Haiying

(Research Institute of Petroleum Processing, SINOPEC, Beijing, 100083, China)

Abstract The article introduced the concept of intelligent chromatogram software and compared the difference between conventional analytic software and intelligent software. The necessity of development from conventional data processing to intelligent software for some normal analytic methods was presented. The development ideas and application experience of this kind of software was discussed with actual application cases in detailed hydrocarbon analysis of gasoline, simulated distillation, and determination of the refinery gases.

Key words Intellectualization; software; Gas chromatography

数据处理是石油化工领域气相色谱分析过程的重要环节之一,色谱分析软件就是为此而开发并应用的。由于历史的原因,现有色谱分析软件的主要功能就是对色谱分析结果按特定方法的要求进行定性或定量的处理,但对于处理结果的合理性与否,或者所用的处理方法是否适合特定的样品分析结果,则尚需专业人员的判断。这种人工干预的结果,一方面使得实际样品的分析效率受到了限制,另一方面,方法应用的成本和效果也因需要具有一定专业水准的分析人员而大打折扣。因此,对一些成熟的、常规的色谱分析方法应用软件,使其向无人工干预的智能型和标准化方向发展就成了一个值得追求的目标了。

人工智能^[1]是一门研究如何模拟人类的思维模式,在机器上实现某些思维或脑力劳动的自动化的方法与技术,是利用计算机模拟、延伸和扩展人的智能的一门技术。从软件程序的内容上来看,它与已往的科学计算不同,更多的是基于规则的一系列逻辑判断,因此,特别适合对现有色谱软件人工干预部分的替换。从原理上来说,如果能将色谱专家的知识 and 经验提取出来并在计算机上实现的话,上述开发无人工干预的自动化应用软件的目标就可能成为现实。

石油化工科学研究院从2005年开始就展开了这方面的应用探索,并取得了一系列的应用成果^[2]。伴随探索应用的过程,我们也逐渐意识到,这种应用技术与传统分析方法嫁接后,带来的可能不仅是方法应用时的便利,从某种意义上说,也许还意味着分析方法自身的革命性进步。与现有需要人工部分干预的色谱分析方法相比,一个完全脱离人工干预的自动化分析方法的应用,除了给应用单位带来方法标准化和人工成本降低、测试效率提升的现实受益之外,还重新定义了商品化分析方法应用软件和专用分析系统的标准,并为新分析方法的应用提供了新的推广模式。因此,对这样一种技术的引入给予重视是十分必要的。本文拟结合实际应用案例,介绍一下智能化色谱分析应用软件开发的思路,以及人工智能技术在这一领域应用的特点,以期色谱分析方法向自动化方向的应用模式转变提供一些可资借鉴的经验。

1 人工智能色谱软件概述

现有色谱分析软件的应用通常主要集中于数值计算,它是一种完全基于算法的程序设计,可将其称为

收稿日期:2007-12-20

作者简介:王亚敏,硕士,高级工程师,主要从事专用分析软件的开发和炼厂气组成分析方法的研究。

普通分析软件。与之相比,智能化软件所包含的范围更广,以色谱分析为例,将涉及到对谱图的理解和概念的处理等,是结合了算法、分析规则和经验公式的程序设计。通过以下的分析比较,可以帮助我们更好地理解二者的差异:

1) 普通软件在解决某类问题时必须要解决该类问题中的每个问题,而智能软件只求解其中被认为合理的或日常可能碰到的问题;

2) 普通软件对问题的解答是基于一组固定算法,只要按步骤执行,肯定能得到答案。而智能软件含有不存在算法的非数学过程,不能保证对每个问题都有精确解。但若某类问题中包含只有抽象思维才能解决的问题,智能软件的优势则会显现。因为它能提供一种找出问题的最优解或合理理解的方向,而普通软件则很可能会毫不费事地给出错误解。

智能软件的设计实现与普通软件有许多共同特征,也要经历需求分析、流程设计、编码实现、调试维护等阶段。但智能软件设计中设计者们还要解决知识获取、表示和推理的问题,这种知识通常涉及对隐含于大量难于精确表征和不断变化的信息之中的规则的总结与提取。

知识获取是设计智能软件的关键问题,是知识表示和推理的前提条件。知识获取途径有两个层次:首先是程序设计者对相关问题的学习和理解;其次,需要应用领域专家或设计者将具体问题的解决方案整理成条理化的文档说明,从有限个能体现某种概念的个别实例归纳出一个适用于普遍实例的概念。

知识表示是将所获取的对于普遍实例的描述予以量化,进而形成一组规则的过程。知识的表示方法很多,较常用的是 IF—THEN 形式的规则,表示为:

RULE(i) : IF Condition THEN Conclusion

其中:RULE(i) 称为规则集中第 I 条规则;Condition 是第 I 条规则条件部分,它可能与知识集中的一个或多个知识相匹配;Conclusion 是第 I 条规则结论部分。

知识推理是智能软件的核心问题。推理过程是对问题的求解策略,是使问题从初始状态转移到目标状态的方法和途径。它可以被看成是一张问题调查表,表中的问题按照前述的知识表示以一定顺序分组,然后可在任何给定的阶段,规定条件分支,这些条件分支执行与否取决于对先前问题的回答。色谱分析中所遇到的数据处理问题是多种多样的,分析专家对问题的解决方式,是因事甚至因人而定,即使对于同一问题也可能存在不同的处理方法。但无论采用何种策略,路径的长短和推理的有效性是衡量推理过程优劣的两个指标。

需要指出,目前智能软件进行知识推理的过程是现代科学仪器 2008 4

依顺序完成的,这一点受限于计算机本身的结构,它不具备人脑所具有的对信息处理的交互性和并行性。最后一点,既然智能软件也是计算机程序,那么,任何一种编程语言均可被使用,只是所用语言不同,软件开发的难易程度不尽相同。

2 应用案例与讨论

2.1 全自动汽油单体烃分析软件

由于具备系统配置简单,操作简便,信息量大的特点,单柱 GC/FID 是目前应用最为广泛的汽油单体烃分析方法^[3],虽然方法本身并不复杂,但如何进行数据的后期处理却是一个难点。用于这一方法的现有色谱软件,通常仅包括对色谱分析数据按既定规则的定性定量计算,无法对计算结果的合理性或具体计算模式的选择进行判断,因此,它的应用离不开分析专家的干预,也就无法实现自动化的连续运行。如果能解决计算模式的自动选择和对分析结果合理性的自动判别,则该应用方法的全自动化就极易实现,这也是人工智能技术在色谱分析软件方面可以有所作为的地方。下面以应用软件中计算模式自动选择功能所要使用的汽油类型判别过程为例,讨论一下人工智能技术的应用思路。

汽油单体烃分析软件对一个分析样品的计算模式^[4]选择是依据样品的类型来进行的,而对样品类型的确认则是通过对色谱图形的识别来完成。面对不同样式的汽油谱图,分析专家的头脑中会出现以往所积累的一些供模仿用的图像,他们通过快速的判断比较,找出与其记忆中一个最相似的模式作为判断结果,这就是人对于图形的识别能力。我们可以将分析专家的判别过程分解为两个方面,一是对谱图的整体感知,包括谱峰分布区间、密集度、最高响应值等;然后再进一步进行细节判断,如特征组分、特征区域等。分析专家之所以能通过上述信息给出对样品类型的选择答案,是因为在他们的头脑中已预存了各种类型的汽油谱图特征。与此相似,应用一种类似于图形识别的人工智能技术,我们也可以通过软件来实现同样的判别。其简单的流程如图 1 所示。

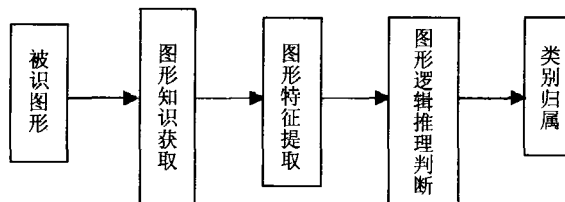


图 1 图形识别过程示意图

图 1 中,图形信息获取就是要充分了解各类汽油

的特性和相互间的差异后,形成概念;图形特征提取是一个知识获取之后的表述过程,通过反复澄清所获得的大量信息,知识变成了程序语言中的各种阈值,而特征提取的全面性和准确性将直接影响判别的正误;图形的推理判断就是求解归类过程,理论上讲,此时会有多个解答路径存在,但从设计的角度应该寻求一种消耗最小代价的解。因此对于一个不错的推理路径而言层次结构非常重要。上述基于专家知识提炼而成的数字化、规则化的谱图识别方式,就是典型的人工智能应用,它可以在很大程度上更有效率地代替专家进行相应的工作。不过需要指出,随着被识对象的变化和对其认识的不断深入,这种智能式判别的规则和推理的过程也可以不断修正和完善,这也是智能软件的特点之一。

2.2 全自动色谱模拟蒸馏分析软件

色谱模拟蒸馏^[5]是采用色谱技术模拟经典的实沸点蒸馏来测定各种石油馏分的馏程,它具有高效、准确、油品用量少等特点。对于此类分析,如何正确和规范化地使用各种分析方法是获得准确结果数据的关键,而在长期的应用过程中,这些关键点的判别均需专业分析人员完成,因此,智能软件的应用目的就是要将这些问题转为由计算机完成。

2.2.1 方法的适用性判断

由于测试对象广,目前的分析技术还无法做到采用单一分析系统完成所有油样的馏程检测,因此在色谱模拟蒸馏的分析中,围绕石油馏分沸程的范围存在着多个标准的分析方法,此时面临的问题就是遇到一个未知油样,如何选择分析模式?在传统的分析过程中,解决这一问题的方法是依靠分析人员的察和问,察主要是观察油品的粘稠度,问是询问油样的大致馏程范围。可否有一种标准方法进行判别呢?答案是肯定的,在全自动模拟蒸馏数据处理系统中,给出了一套以扩展的ASTMD2887^[6]为原型的通用馏程测试方法,此方法测试油样范围宽,包含汽油、煤油、柴油、润滑油、蜡油等,馏程范围从55-720℃,换言之,只要测试样品能完全从色谱柱中馏出,即可采用此方法。

因此,从包含的测试对象范围到操作的简易程度来看,将其作为通用馏程测试方法较为合适。人工判别时,分析人员可通过检查谱图作业流程结束端是否回到恒定基线来实现。结合这一思路,智能程序设计时,可将这一过程量化为判别尾部切片数据前后之间的相对变化率。

同理,由于传统的数据后处理均为离线分析,它也存在数据处理模式的适用性判断,在程序判别时,可主要从每个方法的分析参数入手加以区分。通过统计比较发现,分析时间和数据采集速率是两个重要指标。分析时间的长短隐含了样品在一定测试条件下馏程范围的信息;采集速率则与分析柱的类型相关。例如若是采用ASTM D2887测试条件的样品,根据方法要求,整个分析时间应在40 min以内,采集速率为0.1 Hz,程序设计时,通过对切片数据的简单分析计算和检索,可以很容易地得到上述指标的实际值,若均为真,则可自动转入针对此方法的后续分析,实现了自动化。

2.2.2 色谱模拟蒸馏的全自动分析

从色谱模拟蒸馏的方法原理可知,一个完整的样品馏程分析需要四个步骤:基线分析-正构烷烃混合试样分析-参考油样分析-样品分析。它是一个顺序执行、环环相扣的过程,在传统的分析技术中,对于能否进行下一个步骤均需进行人工判断。这在一定程度上影响了分析方法的规范化,同时也提高了对应用人员的技术水平要求。因此,一些判别技术的引入是解决上述问题的必要条件。以下仅以基线的全自动分析为例进行探讨。

基线分析只有合格与不合格两种结果。合格的结果来自于通盘的考虑,而不合格的结果可能只来自一点,因此若对此类问题进行一个谱图全程式的检索判断,会耗费过多的计算空间和时间,而如果能找到一种方法,确定合理的顺序,即选择那些最有希望的信息加以扩展,那么,搜索效率将会大大提高。下面以宽范围馏分油(扩展的ASTMD2887)的基线判别为例,规则见表1。

表1 扩展ASTMD2887的基线判别规则

节点号	判别准则	判别意义
1	最高响应值与最低响应值的比值	A:是否在进行基线判别 B:是否存在残留量很大的组分
2	最高响应值出现的区域	基线的最高响应值全部出现在特定区间,即温度升到最高点或其后的高温段
3	基线的起始阶段区域面积占总面积的百分比	基线应平稳,不应有过多的残留
4	相邻特定时间段内,后一段的响应值高于前一段	基线随着程序升温出现漂移阶段,漂移应平滑上升,不应存在过多残留
5	最后的恒温阶段最高响应值与最低响应值的比值	程序升温达到最高点后的高温阶段,基线应平缓,不应有过度的倾斜,以防残留和系统泄露

备注:以上判别方式必须配合固定的分析方法(不含溶剂)。

从上表可知,基线的定义和基线分析目的可被作为启发式的信息,其中每一个节点都是最有希望判断基线是否合格的特定标识,基线谱图随时间变化的各个阶段则蕴涵了方法本身的信息(分析条件)。只有当前节点满足要求,才会去判断下一个节点,在这种搜索模式中,某一类复杂问题被分解成符合特定条件的两个或更多的子问题来解决,问题分解的前提是子问题要比原问题易于解决;如果所有子问题都得到解决,则原问题就解决了。

2.3 全自动炼厂气组成分析中的应用

在快速炼厂气分析系统(如 Agilent3000)中,一个全组分分析仅需 160s,而数据分析可能需要 5min,这在一定程度上降低了仪器性能。智能软件在快速炼厂气分析系统中的应用就是旨在保证精度的前提下,最大限度地提高数据处理速度。

通过分析数据处理过程可以发现,影响数据处理速度的主要因素是保留时间匹配过程,当以保留时间为定性模式时,样品组分含量的差异、柱极性变化和系统稳定性等因素都可能改变组分的保留时间,从而增加了定性过程的周期。在以往的解决方案^[7]中,一是根据不同来源样品建立不同的分析方法,即不同的校正表,以此解决含量变化问题;二是建立一张柱极性变化表,针对由于各种因素导致的保留时间变化,人工选择一组数据进行组分定性。这样的处理方案中,分析人员所做的工作就是进行数据类比,而这恰恰是计算机所擅长的,可以利用化学计量学^[8]的相似性判断来模仿人的类比方式去解决这一问题。具体过程是:建立一个数据池,有目标地收集因不同因素导致的保留时间数据,将其构成一个 $M \times N$ 的阵列(M 表示收集样本的个数, N 表示需要定性的特定组分名称),其中每一个数据代表特定状况下的保留时间。对于未知样品可利用其与数据池中每一个样本的距离为指标进行选择(例如欧氏距离^[8]),计算完成后,将其中距离最小样本定义为与未知样品分析状态最接近的模板,最后采用最近点搜索的方式对未知样品的每一个组分定性。

2.4 几点体会

1) 实际分析方法应用中可能遇到的人工智能问题是多种多样的,但本论文所介绍的专家知识的提取流程是相似的,开发人员应该有对分析方法中可应用人工智能技术解决的问题有良好的洞察力。

2) 在色谱分析领域,智能技术的应用范围主要涉及一些视觉信息的处理,或是并不复杂的过程控制问

题,因此对于一个具体分析项目,高深的算法或数学模式往往不是解决问题的最好手段,而特定领域内的特有技巧,甚至包括一些他们的日常行为规范却可能是最佳方式。

3) 智能软件的智能性同人的智能一样也是分层次的,目前的智能软件多停留在对已知问题的感知和处理上,均不具备自学习能力和自处理能力,因此人工智能技术的应用最好选择那些标准分析方法或已十分成熟的分析项目实施。

3 结 论

人工智能软件开发的基本条件是要拥有特定应用领域的技术专家,优秀的技术专家是决定人工智能应用软件性能优劣的首要条件,也是不可缺少的必要条件;而对专家拥有的专业知识、经验和技巧的挖掘、提取和数字化处理,则是软件开发人员开发这类软件需要特别用心的环节。

人工智能的应用,可以极大改善色谱应用软件的性能,它的应用推广为专家知识的共享提供了良好的平台。

色谱分析软件的自动化应该是未来发展的主要趋势,而全自动色谱应用软件中人工智能技术的应用也将成为越来越普遍的现象,应用了人工智能的全自动色谱分析软件,在给各种分析方法的应用带来便利的同时,也为一些可以借助软件实现的复杂运算或推理性的分析结果引入分析方法创造了条件。

参考文献

- [1] 蔡自兴,徐光佑. 人工智能及其应用. 北京:清华大学出版社,2005, 1-23
- [2] 王亚敏,杨海鹰. 气相色谱专用分析软件开发的相关问题探讨. 现代科学仪器,2006(5),27-29
- [3] 李长秀,刘颖荣,杨海鹰. 气相色谱法测定汽油烃类组成分析技术应用现状与发展. 色谱,2004,22(5):521-527
- [4] 李长秀等,瑞博汽油组成分析软件,国家版权局登记号 2003S24316
- [5] 杨海鹰等编著. 气相色谱在石油化工中的应用. 北京:化学工业出版社,2005,107-119
- [6] ASTM D2887-2006 Boiling Range Distribution of Petroleum Fractions by Gas Chromatography
- [7] 王亚敏,杨海鹰. 多柱并行气相色谱仪分析炼厂气组成的方法. 中国, CN 03147979. 2005
- [8] 刘树深,易忠胜编著. 基础化学计量学. 北京:科学出版社,1999, 127-134