



一种基于TCBR与RBR融合推理的试飞支持信息系统*

An Information System for Sustaining Flight Test Basing on the Fusion Reasoning of TCBR and RBR

王国栋¹ 杨博² 邓志东³ 李明^{1/1} 沈阳飞机设计研究所
2 国家计算机网络应急技术处理协调中心 3 清华大学智能技术与系统国家重点实验室

摘要: 深入研究了基于文本案例的推理(TCBR)、基于规则的推理(RBR)及两者融合的人工智能技术,并首创性地运用于先进战机的试飞支持信息系统,由此构建了先进战机智能化试飞支持信息系统的整体框架。建立了其案例库、规则库和知识库的结构。采用Tomcat7、IK Analyzer、GWT、MySQL等运用软件进行开发。实现了理解问答、智能搜索和案例推理等功能,对于试飞工程师还增加了数据库维护及试飞文档编写助手等功能。运行结果表明:系统整体构架设计合理实用,特别是TCBR和RBR的融合运用使系统体现了接近人类思维特点的智能特征。

Abstract: The deep research is carried through on the artificial intelligence about the reasoning based textual cases (TCBR), the reasoning based rules (RBR) and both fusion, and this technique is originally used in the information system sustaining the flight test of advanced fighter aircraft (ISSFTAFA). Moreover, the overall framework of ISSFTAFA is constructed. The case-base, the rule-base and the knowledge-base of ISSFTAFA are set up. The system has been developed utilizing some applied software such as Tomcat7, IK Analyzer, GWT, MySQL and the like, and it realizes the functions such as comprehension-interlocution, intelligent searching and case-reasoning. Additionally, it also adds the functions of database maintenance and writing assistant of documents for the engineers of flight test. The operation results indicate that the design of the overall framework is reasonable and practical. Especially, it makes the system represent some intelligitized characters approaching human thought traits to apply TCBR and RBR synthetically.

关键词: 试飞支持信息系统; 基于文本案例的推理; 基于规则的推理; 融合推理; 实现

Keywords: information system sustaining the flight test, TCBR, RBR, fusion reasoning, realization

0 引言

先进军用飞机的设计定型/鉴定离不开飞行试验,而飞行试验则是飞机在真实环境条件下进行的各种试验。飞机的飞行试验既是探索航空技术未知领域的一门综合性学科,也是其发展和鉴定或定型的重要手段。特别是在研发新型飞机的过程中,飞行试验更是不可或缺,其作用举足轻重。就先进战机而言,

其飞行试验的内容通常包括飞机基本飞行性能、结构动力学特性、振动/冲击/噪声/温度环境、结构强度/载荷、机电系统、飞行品质、飞行控制与操纵系统、航电系统、武器系统等等。而对于舰载飞机而言,通常还需要从母舰甲板上滑跃起飞和进行拦阻着舰,且其飞行试验还要经历常规陆基、陆基试验场和舰基等环节,因此,无论是试飞内容还是试飞

风险及复杂程度都是陆基飞机无可比拟的。如此众多、甚至复杂的试飞内容需要完成,必然涉及大量、甚至庞杂的领域及非领域知识。

将试飞涉及的常识与众多领域知识,如试飞任务剖面、驾驶员操作要领、相关应急处置、使用限制、特征曲线或特征数据等进行归类并以一定的形式进行表示,即所谓的人工智能的知识表示,建成专家知识库/案例库;采用相应的软件工具进行可视化编程,对该专家知识库/

* 基金项目:航空科学基金(2010ZD01003)项目资助。

案例库进行管理;再运用基于案例的推理等人工智能的方法使其具有一定的推理或智能搜索等能力。如此构建出的智能化试飞支持信息系统,既可帮助试飞员或试飞工程师从繁重的学习、记忆与理解的负担中解脱出来,又能得心应手地获取完成新任务所需的知识或解决新问题所需的答案或参考。

从国内外现有的资料来看,智能化的试飞支持信息系统目前尚未有任何报道。究其原因,可能有两方面:一是出于保密的考虑;二是将人工智能的一些方法恰到好处地运用于试飞支持信息系统是一项非常困难的事,不同于决策支持系统,更多的需要是对问题的精确解答。

但随着人工智能技术的不断发展,诸如基于案例的推理和基于规则的推理(专家系统)等人工智能的技术方法可以用来解决更多、更实用的问题。

基于案例的推理(Case-Based Reasoning, CBR)^[1-3]是一种新兴的机器学习和推理方法。在基于案例推理中把当前所面临的问题或情况称为目标案例(target case),而把记忆的问题或情况称为源案例(base case)。简言之,基于案例推理就是由目标案例的提示而获得记忆中的源案例,并由源案例来指导目标案例求解的一种策略。推理循环由4个基本过程组成,即4R循环——Retrieve、Reuse、Revise、Retain,分别对应于案例的检索、重用、修订和保存(学习)。基于文本案例的推理(Textual Case-based Reasoning, TCBR)^[4-5]技术是基于案例推理研究的一个重要分支,它是为了处理以非结构化数据形式存储的文本知识而产生的。人类数千年的知识、经验积累,大部分是以文本形式存储的,如何通过对保存在文本中的知识进行推理,帮助人们决策,是CBR长

期以来要面对和解决的问题,此种情况下,TCBR便应运而生。

基于规则的推理(Rule-Based Reasoning, RBR)^[6-7]是在掌握领域基础知识的前提下,把领域专家的知识或经验转化为规则,将相关领域的问题与解答联系起来,通过启发式经验知识进行推理。具有明确的前提,能得到确定的结果。基于规则的推理实际上就是专家系统,是人工智能应用最为成熟的一个领域,也是人工智能应用研究中一种最早、最有成效的技术系统,自1965年第一个专家系统DENDRAL在美国斯坦福大学问世以来,历经几十年的研究开发,已经成功应用于医疗诊断、化学分析、军事决策等众多领域。

如何将基于文本案例的推理与基于规则(自然语言描述)的推理技术加以融合并运用到试飞支持信息系统之中,是一个崭新的课题,没有成熟的运用经验可以借鉴,需要全新的创造。

本文首先构建了智能化试飞支持信息系统的整体构架,然后阐明了

TCBR 与RBR的融合运用及智能化试飞支持信息系统的软件实现,最后给出了结论。

1 智能化试飞支持信息系统的整体构架

建设先进战机的智能化试飞支持信息系统,其主要目的是为试飞员及试飞工程师等主要用户提供方便、快捷的知识服务,包括对试飞任务剖面、驾驶员操作要领、相关应急处置、使用限制、特征曲线或特征数据等知识或案例的查询或智能推理等。单机版硬件可以是便携式电脑,也可以是掌上电脑;网络版则包括服务器、客户端、加密机、防火墙、基础网络设施、有线专用网络或无线军用卫星通信信道等。而软件系统的核心则是知识库/数据库/案例库及推理机制。

无论是单机版还是网络版,智能化试飞支持信息系统都应包括知识库/案例库/规则库、知识/案例/规则获取模块、基于案例的推理/智能搜索等功能模块、必要的理解/学习模块以及用户界面。

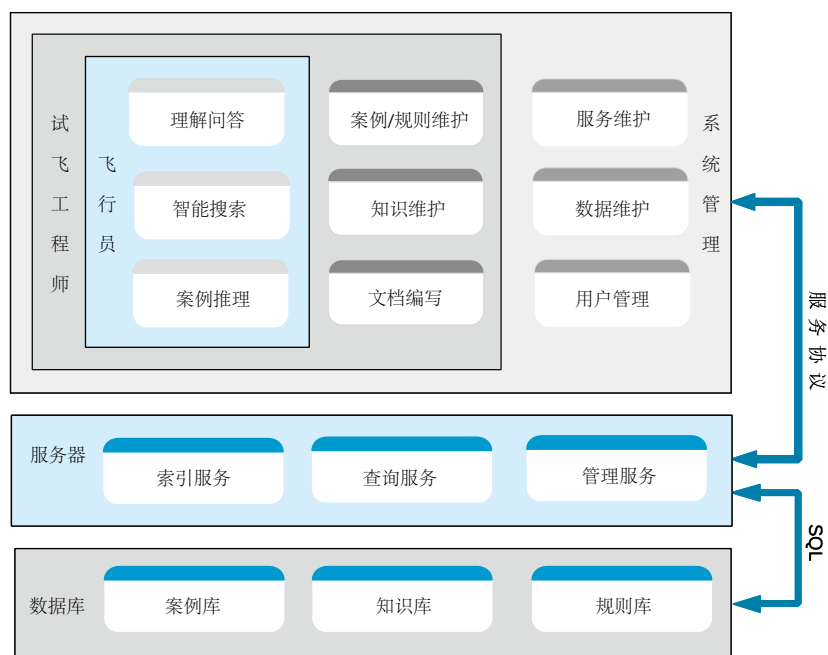


图1 智能化试飞支持信息系统整体构架



从发展的趋势和使用的场合来讲, Browser/Server(B/S, 浏览器/服务器)架构无疑成为了知识系统实现的最佳架构选择。B/S是随着Internet技术的兴起, 对C/S(Client/Server)结构的一种变化或者改进的结构。试飞支持信息系统可采用B/S架构来进行设计, 其主要组成部分如图1所示。

服务器是向用户提供应用的接口。在试飞支持信息系统中包括案例索引、案例查询以及数据库管理等服务, 是用户与数据库内容进行交互的中枢。

在B/S架构中, 客户端通常是指浏览器(Browser)。浏览器是万维网(Web)服务的客户端浏览程序, 可向万维网(Web)服务器发送各种请求, 并对从服务器发来的超文本信息和各种多媒体数据格式进行解释、显示和播放。客户端上提供给飞行员使用的功能包括: 理解问答、智能搜索(又分案例搜索与知识搜索)及案例推理; 对试飞工程师而言, 功能上还包括案例/规则/知识维护及文档编写; 对系统管理员而言, 还包括服务维护、数据维护和用户管理等功能。

数据库用来存放试飞支持信息系统的案例、知识以及推理规则等, 是试飞支持信息系统的数据核心。在实现中通常使用成熟的DBMA来进行存储, 如MySQL、Oracle等。在B/S架构中, 浏览器(Browser)和服务端(Server)通过国际上的一些标准协议进行通信, 而服务端又和数据库(DataBase)通过某种特定的协议进行交互。而对于后者, 为了保证数据的安全, 其交互协议可以是符合标准的, 也可以是自定义的。

2 案例库/规则库/知识库

案例库存放了在试飞任务中可能出现的各种案例, 包括试飞任务剖面、飞行员操纵要领、应急程序、使用限制等各

种案例, 是软件功能中智能搜索与案例推理的主要数据来源。

在实际使用中, 案例分为两种, 一种是可以由很多属性来描述的一个案例, 比如旅游路线案例中, 可以使用天数、费用以及路线上的景点个数来描述; 一种是使用自然语言描述的案例, 比如一个事件的陈述等。在试飞支持信息系统中, 绝大部分案例都需要使用自然语言加以描述, 如果要按照属性描述来整理的话, 需要耗费大量的人力物力, 而且整理出来的属性表会十分庞大, 在计算上会产生高维灾难。

自然语言描述的案例采用一种半结构化的组织形式, 称之为范式。在系统中的案例使用的范式如下:

$$\text{Case} := \langle C1, C2, N, F, T, O \rangle \quad (1)$$

其中C1、C2表示案例所属的一级和二级类别, N表示案例名称, F表示案例特征, T表示处置方法, O表示其他一些附加说明。使用该范式表示的案例称为半结构化的案例, 范式中每一个元素的表示形式是使用自然语言描述的, 使用自然语言的处理方式从中提取相应的特征。

以“发动机舱失火”案例为例, 用该范式表达的案例结构如下:

C1	应急程序
C2	动力装置故障与处置
N	发动机舱失火
F	(1) 仪表板上部的红色紧急信号灯(“火警”)灯信号显亮(闪亮状态);
T	在停机坪启动发动机之后, 在起飞前滑行中(着陆之后的滑行中雷同): (1) 将发动机油门杆置于“停车”位置, 并采取措施使飞机停止运动;
O	灭火以后, 禁止启动着火过的发动机。在可能继续飞行时, 在就近机场着陆。

规则库主要存放了构成专家系统推理基础的产生式规则, 而规则是由专家总结出来的基于因果关系的产生式。

规则库的主要作用是为基于规则的推理提供规则支持, 其中每一条规则是在对案例的特征进行深入分析后得出的, 通常使用精炼的、简短的语言来阐述。在本系统中, 规则的制定是和案例密不可分的, 可以说每一条规则就是案例特征描述的“浓缩版”。

规则库作为系统知识推理的另一条途径, 与案例推理是相辅相成、密不可分的。由于案例库的表示是使用自然语言进行描述的, 因此其中必然包含了很多与实际特征并不相关但是有需要保证语言连贯性的某些冗余词汇, 比如介词、连词等等, 而且自然语言的描述带有一定的不确定性, 有些字词的概念是模糊的、甚至是错误的。而目前自然语言处理的研究成果还不足以将这些干扰因素消除, 因此将专家总结的案例规则作为系统案例推理的数据资源是十分必要的。

由于案例的表述采用了基于自然语言描述的半结构范式, 因此规则的制定和每一个案例密切相关。换言之, 规则是使用高度浓缩的自然语言描述一个案例的, 如“信号灯闪亮”、“飞机尾部冒烟”等等, 因此这里采用基于自然语言的产生式规则如下:

$$\text{IF } (p1 \text{ AND } p2 \text{ AND } \dots \text{ AND } pm) \text{ THEN } Ci \quad (2)$$

其中, p_i 表示用自然语言描述的某种特征, 如“信号灯闪亮”等, 与案例的特征描述所不同的是, p_i 的自然语言描述通常很精炼。 C_i 是该条规则对应的案例。

以“发动机舱失火”案例为例, 用该范式表达的规则如下:

IF (红色紧急信号灯显亮 AND 语音提示左发动机火警 AND 尾部冒烟) THEN 发动机舱失火

知识库是用来存放与试飞支持相关的结构化信息或数据。知识库中知识的结构化形式可以参考案例的范式, 表征如下:

Knowledge := <T, S, C, A1, A2...> (3)

其中 T表示知识的标题, S表示知识类别, C表示知识的内容, A1、A2...表示其它一些附加说明或附件。知识条目以自然语言描述为主, 相关的图表可以放在附件里。

3 TCBR与RBR的融合运用

TCBR系统的目标是检索与解决文本描述问题最相似的文本案例。然而, TCBR处理的是非结构化的文档, 如何进行案例的自动提取、改编和学习, 是一个较为困难的问题。其中, 不可避免地要涉及到自然语言处理(Natural Language Processing, NLP)问题, 如对用户输入的自然语言进行分词等。目前, 可选用的主流分词器有: Paoding(“庖丁解牛”)、Imdict(使用ICTCLAS HHMM隐马尔科夫模型)、Mmseg4j(使用MMSeg算法实现)、IK(使用正向迭代最细粒度切分算法、多子处理器分析模式)等。

基于规则的推理(RBR)自身存在许多缺点, 如规则间的互相关系不明确, 知识整体形象难以把握, 处理效率低, 推理缺乏灵活性, 容易出现控制饱和问题, 与真正专家的知识结构不一致等。为弥

补RBR的不足, 充分发挥TCBR与RBR各自的优势, 这里考虑采用TCBR与RBR融合推理的技术。其推理结构见图2。

当用户输入问题(文本或字段的组合形式)后, 系统作为新的文本/案例, 一方面转化为规则并与规则库匹配, 另一方面启动基于案例的推理, 并把RBR与TCBR给出的解对于用户输入问题的相似度放在一起同时进行比较, 相似度最大者即为建议解。如果该文本/案例与案例库的所有案例相似度小于界定阈值, 用户可以对其进行修正或确认后存入案例库作为新的案例(文本)。

文本/案例的相似度计算方法不一而足。国内研究者提出的有: 基于向量空间模型的TF-IDF(Term Frequency Inverted Document Frequency)方法, 隐性语义标引, 基于海明距离的文本相似度计算方法, 基于属性论的文本相似度计算方法, 基于语义理解的相似度计算方法, 基于词形词序的方法等^[8-9]。国外研究者提出的包括: Solr(一个高性能、采用Java5开发、基于Lucene的全文搜索服务器)中的DefaultSimilarityFactory(默认相似度算法)、BM25Similarity Factory、DFR Similarity Factory、IBSimilarity Factory、LM

Dirichlet Similarity Factory、LMJelinek Mercer Similarity Factory等相似度算法^[10-13]。这里采用综合方法。

综合方法就是以上两种或多种方法综合起来运用的方法。即有:

$$Sim_g(S_1, S_2) = \sum_{i=1}^N k_i Sim_i(S_1, S_2) \quad (4)$$

式中, $k_i \in (0, 1)$, $\sum_{i=1}^N k_i = 1$; $Sim_g(S_1, S_2)$: 为两个文本/案例的总相似度; $Sim_i(S_1, S_2)$, $(i=1, \dots, N)$ 分别对应选定的相似度算法。

4 智能化试飞支持信息系统的实现

智能化试飞支持信息系统的主要功能包括: 理解问答、智能搜索、案例推理等部分。对于试飞工程师, 还增加数据库维护功能和试飞文档编写功能。

如前所述, 试飞支持信息系统构架上分为服务器、客户端和数据库三部分。Web服务器采用Tomcat7服务器(内含Solr搜索服务器); 采用IK Analyzer来搭建分词模型。服务端和客户端选用Java语言来编写, 一则可以方便的将程序移植到不同的PC终端甚至是移动终端上, 二则可以在性能上为试飞支持信息系统提供安全的、鲁棒的实现保证。采用GWT(Google Web Toolkit)来开发客户端页面。数据库采用了MySQL数据库。对Tomcat7服务器、Solr服务、IK Analyzer进行必要的配置。采用SQL语言建立数据库和数据表。对Solr MultiCore、Solr字段、Solr连接数据库进行配置并构建Solr索引来实现服务端与数据库的交互。在TCP/IP协议的支持下, 通过实现GWT RPC的调用、Solr Java Client来实现服务器与客户端交互。开发的试飞支持信息系统功能界面如图3所示。

在功能上实现了理解问答、智能搜索(案例搜索、知识搜索)和案例推理。如当选择“案例推理”功能界面, 用户输入:

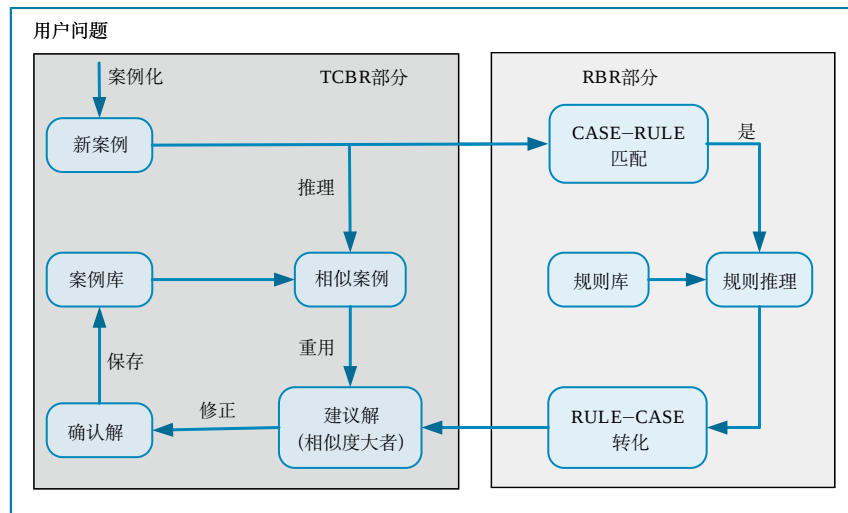


图2 规则/案例融合推理结构



图3 试飞支持信息系统的主要功能界面

“红色紧急信号灯显亮”“发动机火警”，点击“？”(或按“回车”键)则推理返回结果将“发动机舱失火”放在最上端(按相似度大小排列，相似度大者排在前面)。

试飞工程师和系统管理员可以很方便地采用MySQL Workbench来管理数据库。

5 结束语

本文提出了智能化试飞支持信息系统的整体构架，建立了其案例库、知识库和规则库。基于TCBR、RBR等人工智能技术，提出了一种新的TCBR与RBR融合运用的推理技术，设计了文本/案例相似度计算的综合方法。在Tomcat7(含Solr)、IK Analyzer、GWT、MySQL等应用软件框架的基础之上，开发了智能化试飞支持信息系统，验证了基于自然语言的理解问答、智能搜索和案例推理等功能。运行结果表明：本文提出的智能化试飞支持信息系统的整体构架(即服务器端、客户端、数据库三层结构)是合理实用的；将TCBR和RBR进行融合运用体现了接近人类思维特点的智能化特征，尤其是文本/案例之间相似度的计算方法，采用文中提出的综合计算方法，使推理结果更接近于真实。但如何实现语义的自动排歧、怎样处理未登录词或识别专有词语(如“FWS”代表“涡扇”)、如何通用化“理解问答”和“规则推理”机制等

都是需要进一步研究的内容。

AST

参考文献

- [1] 杨健,赵秦怡. 基于案例的推理技术研究进展及运用[J]. 计算机工程与设计, 2008, 29(3):710-712, 721
- [2] 刘冀琼,李兴国,顾东晓,冯帅. 基于案例推理的ISP知识重用方法[J]. 计算机工程, 2010, 36(23):36-39
- [3] 李珺,潘启树,洪炳镭. 一种基于案例推理的多agent强化学习方法研究[J]. 机器人, 2009, 31(4):320-326
- [4] 杨健,张晓玲,周少云. 基于文本案例的推理技术综述[J]. 计算机应用与软件, 2008, 25(7):228-229, 251
- [5] 张文明. 文本案例知识库构建的关键技术研究[D]. 西安:西北大学, 2010
- [6] 江志农,王慧,魏中青. 基于案例与规则推理的故障诊断专家系统[J]. 计算机工程, 2011, 37(1):238-240, 243
- [7] 史忠植. 高级人工智能(第二版)[M]. 北京:科学出版社, 2006
- [8] 沈斌. 基于分词的中文文本相似度计算研究[D]. 硕士学位论文. 天津:天津财经大学, 2006.
- [9] 唐琦. 基于语义分析的句子相似度计算研究[D]. 硕士学位论文. 北京:华北电力大学, 2008.
- [10] Stephen Robertson, Hugo Zaragoza. The Probabilistic Relevance

Framework:BM25 and Beyond [J]. Foundations and Trends in Information Retrieval, 2009, 3 (4):333-389

[11] Glanni Amati, Cornelis Joost Van Rijsbergen. Probabilistic Models of Information Retrieval Based on Measuring the Divergence from Randomness [J]. ACM Transaction on Information Systems, 2002, 20 (4):357-389

[12] Stéphane Clinchant and Eric Gaussier. Information-based models for ad hoc IR[C]. Proceeding of the 33rd international ACM SIGIR conference on Research and development in information retrieval (SIGIR '10). ACM, New York, NY, USA, 2010, 234-241

[13] Chengxiang Zhai and John Lafferty. A study of smoothing methods for language models applied to Ad Hoc information retrieval[C]. Proceedings of the 24th annual international ACM SIGIR conference on Research and development in information retrieval (SIGIR '01). ACM, New York, NY, USA, 2001, 334-342.

作者简介

王国栋, 博士, 高级工程师, 主要研究方向为人工智能技术及其在先进战机上的运用。

杨博, 博士, 工程师, 主要研究方向: 人工智能、现代计算机技术及运用。

邓志东, 博士(后), 教授, 主要研究方向: 人工智能、智能控制技术及其运用。

李明, 中国工程院院士, 研究员, 主要研究方向: 先进飞机设计理论、自动控制及其运用。