# 第一章 绪论

## 1.1课题研究背景和意义

本课题来源于工信部民机专项《航空发动机电子控制系统适航审定关键技术》子项目《航空发动机控制系统软件和电子硬件适航审定方法与流程研究》，主要针对大型涡扇发动机电子控制系统软件的适航审定过程中对在适航审定过程中软件需求的可追踪性目标要求的问题，开展相关的基础技术研究。

适航从字面理解即适合飞行，在中国民用航空局适航司颁发的《中国民用航空器适航管理》中的定义是，是指民用航空器，包括其子系统整体性能和操纵性能在预期的环境和使用限制下的安全性和物理完整性的一种固有品质[1]。我国虽然已经具备自主研制航空发动机的能力，但是在对航空发动机适航审定过程中，仍在面临一些困难：一方面，由于我国航空发动机研制技术相对国际领先水平仍有较大差距，与之配套的适航审定经验不足；另一方面，对航空发动机控制系统的适航审定是的一个重要步骤，这也是未来通过适航审定、取得型号合格证的基础，更是产品进入市场、取得商业成功、实现国家战略的基础。

航空发动机控制系统是保证发动机稳定工作、安全可靠的重要保证 [2]，目前其控制系统中主要采用的FADEC。作为机载软件，在对其的适航审定过程中，一个重要的参考是2011年由航空无线电技术委员会（RTCA）在DO-178B标准的基础上提出的DO-178C，即《机载系统和设备合格审定中的软件考虑》。在我国航空工业中，DO-178C适航符合性标准被引入适航审定过程，如西安航空计算技术研究所(中国航空工业第631研究所)，中国航空发动机控制系统研究所（614 所）的一些项目。DO-178B/C 面向适航审定的过程和目标，明确了研制满足适航性的软件的目标，但是并没有严格约束开发流程，以可协调的方式为开发适航软件提供了指南。在软件研制的生命周期中，机载软件的研制单位要求对其研制的软件产品及其开发过程按照 DO-178B/C 的要求进行内部评审，以发现不符合 DO-178B/C 目标的问题并及时解决。

本论文主要针对在DO-178B/C标准所规定的对适航软件开发过程中，软件需求间可追溯性应该达到的目标和适航审定过程中软件需求追踪链接的构建的相关方法进行了研究，DO-178B/C在所规定的目标中明确指出：1）高层需求可追踪到系统需求；2）低层需求可追踪到高层需求；3）源代码可追踪到低层需求[3]。

广义的软件需求追踪除了系统需求，高层需求及低层需求之间的追踪关系还包括如测试用例，用户样例与原代码等的追踪关系等，用于辅助验证软件的可追溯性。Center of Excellence for Software and System Traceability (CoEST)把软件可追踪性定义为：能够将任何唯一可识别的软件工程制品与其他软件制品关联起来，并在一段时间内维护所需的链接，并使用这些追踪来解决产品及其开发过程前向后向追踪的问题。它是评估软件和系统，尤其是安全攸关软件系统的一个重要的评价指标。完整和准确的可追溯性链接可以确保在变更需求时对每个相关的元素都进行了相关操作，并确保每一个需求都得到实现[5]。在实际的软件开发过程中，常常使用需求追踪矩阵，实体联系模型和追踪图[6]等静态的方式来维持需求间的追踪关，IBM Rational DOORS、IBM Rational Requisite Pro和青铜器RDM等工具可以用来辅助需求的管理。随着软件规模和周期的增长，创建和维护需求追踪不仅需要较高的成本，而且人工判断容易出错，耗时长。而这恰恰是动态建立需求追踪链的方法的优势。动态需求追踪中比较常用的思想是，分析软件制品的相关文件，使用信息检索中的相关技术，如向量空间模型（Vector Space Model，VSM）、隐语义模型（Latent Semantic Indexing，LSI）和主题模型（Topic Model）[7]等方法来计算软件制品间的相似度，然后通过设置相似度阙值作为需求追踪判断条件的依据。他们有一个共同的缺点是，他们把词汇当作最小独立的“单元”，导致这些方法往往会面临“词汇鸿沟”的问题。

词嵌入（word embeddings）是近年来自然语言处理领域的一个热门分支，它可以通过使用神经网络模型对大量无标记的数据进行训练来得到，这种技术的一个重要特性是携带了语义特性，在向量空间距离相近的词语具有相似的语义，在文本相似度计算上具有较好的效果。同时，软件工程领域的相关问题也开始引入机器学习方法，其输入是将文本的不同特征的组合，这种方法也为处理动态需求追踪相关的问题提供了借鉴思路。

综上所述，可追溯性作为DO-178B/C所要求必须实现的目标，在航空发动机控制系统的开发和适航审定过程中有着重要作用。因此，研究适航领域的需求可追踪算法来辅助审查人员半自动或自动化的构建需求间的追踪链，能极大减轻适航审查人员的工作负荷，减少出错率，对航空发动机系统的适航审定有重要意义。

## 1.2国内外研究现状

### 1.2.1 FADEC系统软件概述

从上世纪70年代FADEC开始出现，国外已经在军用发动机F100和民用发动机JT9D上对系统的性能、使用和维护指标进行了验证。经过30多年的发展，FADEC系统已经成为目前航空发动机的主要控制系统[8]，随着我国成功研制出某型号发动机的控制系统，说明我国在发动机控制系统的研制上具有了一定实力。

目前FADEC软件系统主要包括信号采集和处理模块，核心机控制模块，系统监控模块，信号输出模块，冗余管理模块和支撑软件模块等六部分。其模块图如图1所示。



图 1 FADEC软件系统功能模块图

可以看到，FADEC软件系统由多个子模块构成，涉及到数据的采集和处理，系统控制等等方面的功能，软件系统复杂。随着我国航空领域技术的不断发展，对航空发动机的要求也越来越高，随之而来的是航空发动机控制系统软件的需求越来越复杂。

### 1.2.2 软件需求追踪技术

可追溯性作为航空发动机适航符合性审查的目标之一，其重要性早已软件业界认可。软件能力成熟度集成模型(Capability Maturity Model Integration, CMMI)对需求管理提出的明确的指导性的方法，目前已逐渐被软件工程领域所采用，需求追踪作为软件需求管理的重要组成部分，其重要性也是不言而喻。

需求追踪技术主要分为两种类型：静态需求追踪技术和动态需求追踪技术。静态需求追踪技术通常是由手工构建和维护追踪链，常用的方法有：需求追踪矩阵、交叉引用和追踪图。动态追踪技术是为了以自动或半自动化技术帮助开发人员建立和维护可追踪链，这个过程中各种软件制品被看作静态文本文件，根据文本间的关系来构建可追踪链。

#### 1.2.1.1 静态需求追踪技术

静态需求追踪方法优点易于构建，是传统需求追踪中主要使用的方法，但由于只能静态构造，在大型项目或软件需求变更频繁的项目中，这种方法却难以维护和扩展。

需求追踪矩阵（Requirement Traceability Matrix，RTM）是展现和表示需求间追踪信息的最简单方法，它的主要的表现方式是通过矩阵或者表格来表示两个或者多个软件制品间的追踪关系，需要注意的是，软件需求间存在的关系不仅仅是“一对一”的关系，还有“一对多”，“多对多”的关系。一个软件需求追踪矩阵实例如图2所示。



图 2 需求追踪矩阵示意图

追踪图与追踪矩阵类似，它可以用图形的表现形式把用户自定义的对象和关系，将软件需求、测试用例、及源代码等等软件制品表述出来，这种方法直观，灵活，易于理解。

交叉引用则是通过对相互关联的软件制品（实体），将其中一个实体的关联属性中给出另一个相关实体的引用的方式表示。可以在需求文档之间建立追踪关系，需求文档包括需求规格文档、需求说明文档等，对于发动机控制软件，还有系统需求、高层需求和低层需求等不同文档。这种方法表示追踪关系比较直观，方便实际使用。只适用于需求文档间关系的建立与处理是它的一个主要缺点[9]。

#### 1.2.2.2 动态需求追踪技术

为了解决在人工构建并维护软件需求的可追踪关系时费时费力，容易出错，难以维护的问题，有些学者开始尝试自动化或者半自动话的构建和维护需求间的追踪关系。但直到第13届需求工程(Requirements engineering, RE)大会上，“动态需求追踪”这个概念才首次由J. Cleland-Huang提出。

在构建可追踪链的过程中，学者们从不同角度提出了方法来自动构建需求间的追踪关系，主要有基于事件的追踪技术（Event-Based Traceability，EBT），基于场景的追踪技术（Scenario-Based Traceability，SBT），基于程序运行时（Runtime-Based Tracebility，RBT）及基于信息检索技术(Information-Based，IR)[7]等。由于软件在设计、开发和维护过程中产生的生命周期数据大部分为文本信息，而需求之间的关系可以认为是文本间的关系。在这种情况下，基于信息检索技术的动态的需求追踪算法中有广泛应用，基于信息检索的需求追踪算法经典模型[11]如图3所示。

软件工程师

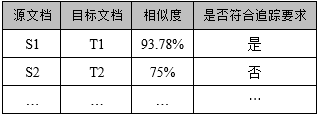
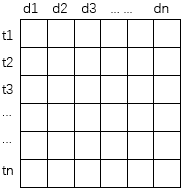
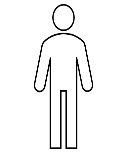
文本向量

结果分析

选择文件

软件制品

文本处理



相似度计算

图 3 基于信息检索的需求追踪算法经典模型

这种方法的主要思想是通过比较文档间的相似度确定需求间可追踪性关系的，从图3可以看到，基于信息检索算法的动态需求追踪算法流程为：首先，需要对软件制品数据作为文档进行预处理，并建立文本的向量表示；然后通过文本相似度算法来计算由软件工程师所选定的查询文件和候选文件间的相似度，最后，通过将文本相似度与事先设定好的阙值去比较，如果相似度足够大，则说明两个文件间具有可追踪关系。显然如何计算文本相似度是需求追踪算法的一个关键问题。

Dag等人[12]首先提出通过使用文本分析的方法来判断需求间的追踪关系，它们通过自定义一些衡量标准，主要通过比较文本间单词的关系计算文本相似度，最后通过调整阙值，来找到最优结果，并取得了一定的结果，并指出了这种方法很有发展前景。同年，G. Antoniol等人[13]使用基于概率模型(Probability Model, PM)和向量空间模型的需求追踪方法为源代码和文档间的追溯性关系，并在自己的数据集上取得了很好的效果。目前的很多基于信息检索的需求追踪算法都是以他们所做的工作为基础的。

Huffman 等人[15]在向量空间模型算法的基础上，有两点改进，一个是手工构建关键词库，另一个是使用简单词典，将一些有关联的词对应起来，起到了扩展查询的作用，最后她们将这种算法应用于NASA的项目，通过与手工和单纯的向量空间算法进行了对比，有较好的效果提升。在这篇文章中她们还整理了NASA项目的需求子集，将高层需求、低层需求和追踪关系整理出来提供给其他研究者继续进行研究。

随后，很多学者通过不同的方法对文本进行表示，从而计算相似度来更快和获取更高质量的需求追踪链的方法，隐语义模型(LSI)和主题模型(LDA)[15,16]被应用于需求追踪，建立需求、软件制品和测试用例之间的跟踪关系，他们通过将具有相同意义的词语映射到同一空间以降低文本向量的维数，以此来降低运算复杂度并获得较高质量的可追踪链。直到现在这些方法已经成为需求追踪领域比较成熟技术[17]。

一些学者在此基础上，进行了改进，将蚁群算法，加入过滤器，本体[18-20]等方法与传统的文本相似度算法相结合，用于提升结果精度。到现在通过改进这种基于统计的信息检索方法所获得的提升已经很小。已经有学者尝试将机器学习和深度学习技术应用于这一领域。

David等人[21]通过用机器学习分类器的方法应用于需求追踪算法，并通过将不同的信息检索算法与多个分类器进行组合，给出了在不同数据集上的最佳实践。它指出可以使用分类器来对追踪链接进行分类，从而不需要在不同的数据集上设置阙值。这说明可以将机器学习方法应用于需求追踪的动态构建中。

而Jin G等人[17]则是将目前流行的word embedddings计算和深度学习技术应用于需求追踪领域。他们将文本用词向量特征表示，并通过循环神经网络（Recurrent Neural Networks，RNN）训练并预测文本间的相似度，他们的实验效果很好，但是它需要一个高质量的领域本体和大量被人工验证过的可追踪链接来进行训练，因此这种方法并不适用与较小的数据集，不具有普遍性。

除此之外，研究者还提供了一些基于信息检索技术的工具，如处理非结构化文本的NETRO，可拓展的实验工具TraceLab，候选链接聚类工具TraCter和支持投票的FaceTrace[22-25]等工具，来帮助学者进行研究。

## 1.3研究目标与研究内容

本文主要对航空发动机控制软件在开发过程中生命周期数据间追踪关系的构建方法展开了研究，根据DO178B/C的中关于需求可追踪性的要求，通过提出的算法自动化的建立软件生命周期数据间的追踪关系，为发动机控制软件的适航审定提供辅助依据。本文的主要工作内容详述如下：

（1）对航空发动机适航理论的研究，DO-178C标准中对软件生命周期的定义与普通软件有所不同，软件研发过程中要求大量的软件生命周期数据，并且对系统和软件的可追溯性具有严格的要求。本文将对适航标准DO-178C及其补充标准进行研究，明确在机载软件的软件生命周期，以及每个过程中应该达到的目标，重点在于标准中与需求跟踪有关的部分，和在软件设计、开发到使用的全过程中需要注意的问题。

（2）需求追踪算法模型构建。动态构建需求追踪链有很多方法。本文提出了一种基于词向量技术的适航领域文本相似度算法，在考虑语义的同时，在算法过程中对词向量缺失词的动态处理；然后结合机器学习领域的学习排序技术进一步提升了需求追踪结果的精度。

（3）算法模型验证。针对提出的需求追踪算法，通过将提出的适航领域文本相似度算法与目前需求追踪中常用的相似度算法进行比较(LSI)，并将最终模型与文献[21]提出的ENRL方法进行对比，对提出的算法模型进行了验证。

（4）最后，基于算法模型，实现了一个需求追踪的原型系统，包括数据交互，数据管理，用户管理及需求可追踪性算法实现模块等。

## 1.4论文组织结构

本文一共分为六个章节，各个章节的编排和主要内容如图4所示。

第一章，绪论。首先给出了课题的来源，介绍了论文的选题背景及意义，同时介绍了当前国内外航空发动机控制系统的发展情况和软件需求追踪领域的研究情况，最后介绍了本文的研究目标和研究内容，并给出整个文章的组织结构。

第二章，相关理论与技术研究。本章围绕航空发动机软件控制系统适航审定所依据的法律，对需求可追溯性目标的要求展开了研究，并对目前基于信息检索的需求跟踪领域常用的关键技术进行了研究，这些技术是本文所提出需求追踪算法的基础。



图 4 本文组织结构图

第三章，面向航空发动机适航的软件需求追踪模型方法。本章通过结合目前较为先进的词嵌入（word embedding）技术和机器学习排序技术，提出了一个需求追踪算法模型，在这个过程中，通过分析适航领域文本特点，提出了一种在使用词向量时动态处理缺失词的方法，并辅助最后文本相似的的计算，接着通过结合学习排序算法，对通过计算相似度得到的候选列表再次处理，从而构建需求间的可追踪链。

第四章，算法模型验证。在本章中，首先为第三章算法模型的两个个阶段分别设置了对比实验来验证各个阶段的有效性，然后与最新提出的需求追踪领域算法进行对比来说明算法的实用性。在本章中还介绍了数据处理，训练词向量及评价指标的实验过程细节。

第五章，需求追踪原型系统的设计与实现。本章按照软件开发过程中的系统需求分析，总体架构设计，数据库设计和具体模块设计与实现依次进行介绍，并在最后给出了该系统的开发环境及工具的详细信息。

最后，总结与展望。本章主要是对全文工作进行总结，并介绍了今后的研究方向和工作重点。