

一种基于装备画像的武器装备数据化方法

刘虎^{1,2}, 刘卫东¹, 杨萍¹

(1. 第二炮兵工程大学, 西安 710025; 2. 中国人民解放军96265部队, 河南 450000)

摘要:为了克服装备数据芜杂性对数据挖掘应用的制约,基于特征工程理论,提出了一种基于装备画像的武器装备数据化方法。基于特征工程中的特征和特征向量的概念,给出了装备标签、元标签、装备画像和视角的概念及其形式化描述;研究了装备标签的建模方法,对装备标签进行分类并分别研究了不同权值的确定方法,给出装备标签创建的一般性原则;分析了装备画像的优势和需要进一步解决的问题。

关键词:数据挖掘;特征工程;装备画像;装备标签

本文引用格式:刘虎,刘卫东,杨萍.一种基于装备画像的武器装备数据化方法[J].兵器装备工程学报,2016(3):59-62.

Citation format:LIU Hu, LIU Wei-dong, YANG Ping. Method of Weapons and Equipment Data Based on Equipment Profile [J]. Journal of Ordnance Equipment Engineering, 2016(3):59-62.

中图分类号:TP311.5

文献标识码:A

文章编号:1006-0707(2016)03-0059-05

Method of Weapons and Equipment Data Based on Equipment Profile

LIU Hu^{1,2}, LIU Wei-dong¹, YANG Ping¹

(1. The Second Artillery Engineering University, Xi'an 710025, China;

2. The No. 96265th Troop of PLA, Henan 450000, China)

Abstract: In order to solve the problem that miscellaneous equipment data restricts the application of data mining, a method of weapons and equipment data based on the theory of feature engineering was presented. Based on the concept of feature and feature vectors, the concept and formal description of equipment tag, meta-tag, equipment profile and angle were given. The method of modeling equipment tag was studied, which includes classification of equipment tag, the technique of determining different weight of equipment tag and the general principle of equipment tag creating. The advantage of equipment profile and some future directions were summarized.

Key words: data mining; feature engineering; equipment profile; equipment tag

数据挖掘技术在民用领域的长足发展给军事研究人员带来启发,利用数据支持装备管理决策和指挥决策展现出更深的潜力,并成为近年来装备管理研究领域的重要方向^[1],随着作战数据工程的建设和完善,更是为基于数据挖掘的装备决策支持从理论走向实践提供了可能^[2,3]。然而,现代化装备的复杂性所带来的数据芜杂性也为挖掘和决策运用带来诸多不便,主要研究体现在3个方面:一是如何有效地组织和规范化数据,以方便数据挖掘算法的实施;二是如何有

效地进行数据挖掘,以得到可用于实践的知识;三是如何有效地展示数据挖掘结果,以保证挖掘结果能更好地指导实践。

用户画像是近几年在社交网络中提到的重要概念^[4-7],随着互联网数据积累,用户画像逐渐成熟,取得非常好的效果,是一种经过实践检验的重要技术。受到用户画像启发,基于特征工程理论,提出装备画像的概念,给出了装备画像、装备标签、元标签和视角的形式化描述,对装备画像中装备

收稿日期:2015-08-10;修回日期:2015-08-30

基金项目:军事学研究生资助课题

作者简介:刘虎(1986—),男,硕士研究生,主要从事军事决策支持、数据挖掘研究。

标签的建模方法进行研究,最后展望了装备画像的应用和下一步需要解决的主要问题。

1 特征工程概述

影响数据挖掘品质的一个因素是让研究人员望洋兴叹的芜杂大数据。怎样对杂乱无章的数据进行整理,使其能够更好地服务于数据挖掘,特征工程(Feature Engineering)的相关研究给人们提供了一个新的思路^[8]。

在机器学习和模式识别领域,通常用特征向量来表征某个个体^[9]。特征向量是个体在多个维度上的观测,一般采用数值表示,对于某些非数值化的特征,通常需要采用一些技术手段将其转化成数值特征,以方便计算机进行处理。本研究涉及的文献,通常对特征向量和特征不做区分,在本文中,特征指个体在单个维度上的观测,特征向量指多个特征的有向集合。

在数据挖掘中,特征工程依据领域知识,将个体若干维度的数据转化为特征向量,从而更好地将数据应用于关联分析、分类、聚类和预测等数据挖掘技术。特征工程研究的内容通常分为特征获取、特征选择、特征创建、特征变换等。

2 装备画像:装备标签的集合

在特征工程中,采用特征表示个体某个维度上的观测,特征是一个数值,虽然有利于计算机进行处理,却不能形象化地展示,为此,提出同时包含语义化特征和数值特征的装备标签的概念。

2.1 装备标签

装备标签是一个同时包含装备特定属性和值的二元组,其形式化描述为:

$$Tag = \langle Name, W \rangle \quad (1)$$

其中:*Name*表示属性名称;*W*表示权值,其类型和取值范围由属性决定,并由下文提到的元标签给予规定。比如,表示维修性的某个装备标签可以表示为〈维修性,2〉。

装备标签可以是对现实装备多种属性的抽象和聚合,也可以是对某种属性的忠实描述。比如一个装备标签可以是〈可靠性,3〉,也可以是〈经度,124.21〉。

装备标签与特征的不同之处在于装备标签同时给出了属性,在*Key-Value*类型的数据库^[10]中,可以进行很好地存储。为了更好地进行人机交互,将使用元标签对属性和值进行格式化。

2.2 元标签

元标签主要描述了装备标签的权值类型、显示方式等,可以理解为“标签的标签”,将其形式化描述为:

$$MetaTag = \langle Name, ViewAngle, WeightType, DisplayMode \rangle \quad (2)$$

其中:*Name*为标签的名称,主要表示该元标签所对应的装备标签;*ViewAngle*为标签的视角,一个标签可以对应多个视角;

*WeightType*为标签权值的数值类型,如布尔值、离散值、连续值等;*DisplayMode*为标签的可视化显示方式,根据标签的意义不同,可以定义多种不同的表示方式,比如对于标签〈维修性,1〉,可以显示为“维修性等级:1”,“维修性非常好”,根据装备标签重要性不同,还可以定义标签在可视化时显示的尺寸。

装备标签只是装备在单个维度上的观测,为了便于在整体上对装备进行描述和处理,进一步提出装备画像的概念。

2.3 装备画像

装备画像是装备标签的集合,可以采用四元组来进行描述:

$$Profile = \langle ID, Name, Label^N, T \rangle \quad (3)$$

其中:*ID*表示装备画像的编号;*Name*表示装备画像的名称;*Label^N*表示装备标签集合;*T*表示形成画像的时间,根据画像形成时间不同可以确定画像的代数,从而有利于进行装备的趋势分析。

装备画像可以容易地转换为形象化文字描述,比如一个装备画像可以转化为以下的文字描述:XX型坦克,作战区域为西安,服役时间9年,射程15 km,累计训练时间为2 736 h,可靠性等级为1级,可靠性很高,维修性等级为3级,维修性一般……。

2.4 视角

为了更有针对性地对装备进行分析,很多时候人们并不需要装备的完整画像,也就是说,只需要一部分装备标签就可以解决问题,这实际上是特征工程中特征选择的问题,基于此,提出视角概念。

视角是装备画像中标签集合的子集,依据标签子集,可以达到解决某领域问题的目的,视角具有与装备画像相似的形式化描述:

$$ViewAngle = \langle Name, Label^M, T \rangle \quad (4)$$

其中 $M \leq N$ 。

3 装备标签的数据建模方法

装备标签是装备画像中的一个重要概念,本节主要研究标签权值的数值类型和不同标签权值的确定方法,并提出标签创建的一般性原则。

3.1 标签权值的数值类型

标签的权值可分为布尔值、离散值、连续值和可枚举值。布尔值类型指非此即彼的数值,也称为二值类型。比如一台装备不可能既具有动力系统不具有动力系统,通常用0表示否定,用1表示肯定。

离散值是指在取值上不连续的值,通常用来表达数量,比如装备含有液压元件的个数为5,其标签形式可以表示为〈液压元件,5〉。

连续值是指在取值上连续的值,比如武器装备的射程、维修工时、服役时间等。

可枚举值实际上是有限个离散取值,在装备标签中,主

要是表示装备的各类等级。比如一台装备的品质等级可分为新品,堪用品,待修和待报废,可以分别使用1,2,3,4表示。可枚举值能够处理为布尔值。比如对于标签〈品质等级,2〉,可以转换为:〈新品,0〉,〈堪用品,0〉,〈待修,0〉,〈待报废,0〉。

3.2 标签分类及权值的确定

根据装备管理的实际,将装备标签分为属性标签、行为标签和动态评价标签,以下将介绍其含义及标签权值的确定方法。

3.2.1 属性标签

属性标签表示了装备固有属性。比如某个标签可能是〈液压元件,20〉,这个标签表明了某台装备安装有20个液压元件。属性标签通常是装备的一些静态属性,一般不随时间的变化而变化,并且其取值通常也很容易得到。

3.2.2 行为标签

为了分析一台装备的维修情况,人们常常会问以下几个问题:这台装备进行过多少次维修?基层级维修、返厂维修和社会化维修分别进行过多少次?最近经常进行何种类型的维修?

首先,次数是分析维修情况的一个重要考虑因素,次数越多,表明影响越大(需求越高);其次,维修工时是衡量维修情况的另一个重要方面,比如一台装备只进行过一次长达半年返厂级维修,进行基层级维修的次数为10次,可每次只维修了一天,显然返厂维修对装备的影响比较大;最后,维修的时间也是一个重要的因素,比如某台装备一年前进行基层级维修的次数较多,而今年以来进行返厂维修的次数较多,显然返厂维修对目前的装备来说比较重要。

根据以上分析,行为类的标签的权值按照(5)式进行确定:

$$W = \frac{1}{1 + a \times e^{-x}} \quad (5)$$

其中: a 是通过训练得到的一个正数; $x = n \sum_{i=1}^n \omega_i t_i$; n 表示某种维修类型总的维修次数; ω_i 表示某次维修活动所在时间窗的重要度,越接近现在,则该重要度越大; t_i 为该时间窗内某次维修的工时,时间窗的长度和重要度可由实际情况确定。

式(5)作为一个基本公式可以得到大多数行为标签的权值,区别只是随着行为特征的不同,对 ω_i 和 t_i 进行不同的取值,比如装备故障类标签“启动失败”可能与持续时间并无关系,这时就可以将 t_i 取为1。

3.2.3 动态评价标签

在对装备的长期使用和管理中,人们已经在很多维度展开了对装备的评价。比如,使用可靠性度量装备完成作战任务的能力,使用维修性武器装备发生故障后修复的难易程度,使用生存能力评价武器装备在外部作用下保持执行规定功能的能力,其他评价指标还有防护能力、侦察能力、抗干扰能力等。在传统的方法中,这些指标通常通过机理性建模或者仿真模拟的方法得到。对于简单的武器装备,机理性建模

是一种比较好的方法,对于复杂装备由于影响因素的增加,给机理性建模带来很大挑战。仿真模拟的方法通常应于装备系统或装备体系,并且具有很高的时间复杂度。同时,机理性建模和仿真模拟的方法通常应用于装备定型和验证方面,也就是说,这两种方法均很少考虑装备随着的使用而带来的能力指标变化。但是,对于一台装备来讲,其各方面的能力往往是随时间变化的,比如过去一年的可靠性与今年的可靠性可能会有很大不同。为了表示装备性能或能力的实时性度量,建立一种动态评价标签。

动态评价标签是为了表示对装备某一方面能力或者性能的动态度量。动态评价标签是对一台装备某方面能力的实时评价,直接关系到管理决策和指挥决策的制订,是装备标签体系中一个很重要的方面。为了有利于下一步进行数据分析和挖掘,采用分级的方法对装备的动态性能进行评价,动态评价标签的权值可以采用枚举值类型。

传统的机理性建模和模拟仿真方法在动态评价标签权值确定时存在局限性,接下来将以装备动态维修性为例介绍一种利用数据挖掘中分类方法确定动态评价标签权值的基本思路^[11]:

步骤1:准备训练数据,包括特征提取和数据预处理。特征提取可以理解为提取与动态维修性相关的数据,动态维修性与装备的类型、元件类型和数量、工作时间等都有很大关系,在实际操作中,可以采用专家研讨的方式对相关因素进行确定。数据预处理主要是进行数据抽取、清理和标准化。事实上,前期在确定属性标签和行为标签的权值时,已经进行了大量的预处理和标准化。

步骤2:选择分类器。分类算法在数据挖掘领域中进行了广泛的研究,像ID3算法、贝叶斯分类、遗传分类算法、神经网络方法等在实际应用中都有非常稳定的表现。

步骤3:训练分类器。根据分类器的不同有不同的训练方法,这里不再赘述。

步骤4:应用分类。这一步是将训练好的分类器应用于新数据,可以得到动态评价标签的权值。

3.3 装备标签创建的原则

装备标签是表示装备某个维度上的特征,是形成装备画像的基础,标签创建的品质关系到装备画像能否准确地表示一台装备,因此,装备标签的创建非常重要,以下总结7条创建装备标签的原则:

1) 相关性。标签之间具有相关性,并非完全孤立的。比如标签“启动失败”与“修理一连”具有相关性,因为车辆爆胎后的承修单位是修理一连。

2) 多视角。标签可对应一个或多个视角,比如维修性在日常管理中关系到维修资源的统筹,在任务规划中关系到保障资源的运用,因此,其既对应管理决策视角,也对应作战决策视角。

3) 定性定量相结合。定性体现在标签的描述上,主要是为了展示;定量体现在标签的权值上,目的是为了下一步进行数据挖掘。

4) 灵活性和自明性。自明性是指标签的描述应达到无需进一步解释的要求,保证快速进行决策。标签的灵活性体现在标签描述形式上的多样化,比如一个标签可以是“射程XXX km”,也可以是“覆盖A国全境”。灵活性是自明性的重要保证。

5) 客观性。标签不等同于指标,指标表达了人们对特定群体的一种期望,比如称A型武器装备的CEP为0.3 m,“0.3”是在装备研制过程中通过试验确定的平均值,但是A型武器装备的个体并不一定每台都达到此要求,有的偏大一些,有的可能偏小一些。标签对应指定的个体,更需要强调客观性。

6) 时变性。时变性是客观性的一个重要体现,这个特性主要体现了装备随时间变化而变化,装备的品质会下降,从而造成某些标签的内容会产生变化,比如可靠性随着时间的变化可能会降低。

7) 冗余性。为了更好地描述一台装备,应当允许适当的冗余,比如在描述装备位置的时候,既需要提供经度、纬度、高程等以实现精确的表达,也应当标识地名等以实现直观地显示。

以上只是创建装备标签的一般性原则,在实际中,还要结合具体问题进行操作。

4 装备画像的应用

装备画像是基于特征工程对装备数据的进一步规范化和标准化的方法,由于植根于数据挖掘,因此可以应用于传统的数据挖掘技术,并由于自身的特点,具备传统基于数据挖掘的分析方法所不具有的优势,本节将对其基本应用进行描述

4.1 装备可视化

装备画像可以通过计算机视觉技术转化为形象化的图像,以某台99G型坦克为例,其可视化图形,如图1所示。

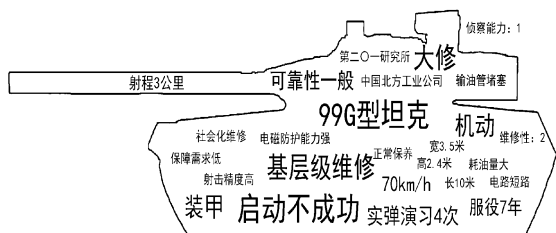


图1 装备画像的可视化

从图1中可以看出,这是为一辆99G型坦克进行的画像,从其中可以很容易地看出坦克的基本参数、维修、保障、执行任务、动态能力水平等情况,该坦克服役7年间进行了4次实弹演习,发生过启动不成功、输油管堵塞和电路短路等故障,但是启动不成功是经常发生的故障,在维修方面,该装备进行过基层级维修和社会化维修,还进行过大修,基层级维修和大修的比重比较高。

4.2 数据分析

由于装备画像具有时间属性和丰富的标签,因此,可以在很多维度上进行数据分析。在时间维度上,可以调用此坦克以前的装备画像,画出“启动不成功”标签权值的趋势图,研究是否可能是由于装备老化或环境变化导致的问题;在同型号坦克的维度上,可以比较其他型号的坦克是否经常出现这类故障,查找是否装备所在单位存在管理上的问题;在生产厂家维度上,可以观察“中国北方工业公司”出产的同类装备是否也经常出现“启动不成功”的故障,从而为采购做决策。

4.3 预测

预测是装备画像比较高级的应用。预测在装备画像的应用可以分为两类:一是服务于装备画像自身,主要是进行标签的构建和更新,比如一些动态评价标签的权值可以通过预测的方法得到;二是服务于装备管理实际,一个典型应用是通过装备历史上的训练、维修、故障、保障等情况,推测装备未来的保障需求。

5 需要研究的主要问题

5.1 装备画像管理体系构建

装备画像是装备标签的集合,而装备标签本质上是数据,怎样对装备画像进行管理和规范是装备画像走向实际应用过程中需要解决的重要课题。本研究提出元标签的概念,实际上可以认为是装备画像与装备仓库之间实现数据交流的接口,也是构建装备管理体系的基础。

5.2 视角的建模方法

本研究提出了视角的概念,但是并未对视角的建模方法进行介绍。事实上,从特征工程的角度来看,视角的建模是特征选择的过程。特征选择是特征工程所需要研究的一大类重要问题,基于特征选择的视角建模方法,是本文下一步研究的重要方向。

5.3 装备群体画像

装备画像是针对单台装备进行的建模,在实际应用中,决策人员很多时候对装备系统更感兴趣,这是因为随着信息化水平的提高,装备之间的耦合性增加,装备系统往往会产生单台装备无法体现的特征,针对整体装备系统进行画像也就具有更深远的意义。

6 结束语

本研究基于特征工程研究了装备画像方法,提出有关的概念并重点对装备标签的构建方法进行了研究,展望了装备画像的应用前景,并指出了下一步需要研究的主要问题。装备画像是数据挖掘在装备管理领域的典型应用,同时该技术也能够有效解决数据冗余性的问题,是在装备管理领域应用数据挖掘技术的重要抓手。通过进一步研究和完善,必将更好地为装备管理决策服务。

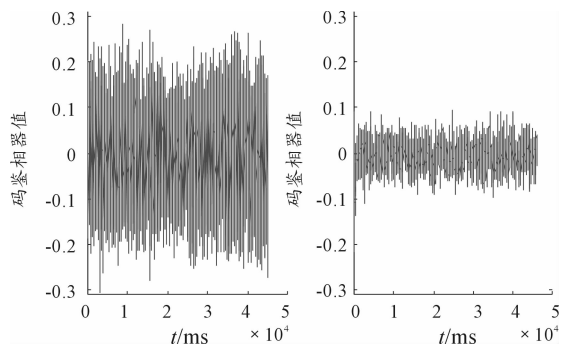


图9 码鉴相器输出

4 结论

本文提出的北斗/惯导深组合捕获和跟踪的控制方法,一方面,利用6 ms并行码相位搜索捕获的方法,能够在捕获过程中规避NH码影响;另一方面,通过中低精度惯导的解算结果推算接收机跟踪环路控制量,在惯导精度不高的情况下实现了环路的持续跟踪。试验表明,该方法能够成功实现北斗卫星的捕获并且提升跟踪环路的性能。

参考文献:

[1] 谭述森. 北斗卫星导航系统的发展与思考[J]. 宇航学报, 2008, 29(2): 391-396.

- [2] 杨洋. GPS/SINS 深组合导航中的关键技术研究[D]. 南京: 南京理工大学, 2013.
- [3] TANG X, FALLETTI E, PRESTI L L. Fast nearly ML estimation of doppler frequency in GNSS signal acquisition process[J]. 2013, 13: 5649-5670.
- [4] ZHAO S, LU M, FENG Z. Implementation and performance assessment of a vector tracking method based on a software GPS receiver[J]. Journal of Navigation, 2011, 64(S1): S151-S161.
- [5] 董俊杰. 基于 Matlab 的北斗二代 B1 频点软件接收机研究与实现[D]. 西安: 长安大学, 2014.
- [6] 陈建卫. 基于 NS210-B 的北斗卫星导航系统 B1 信号的研究[D]. 南京: 南京理工大学, 2014.
- [7] 王新龙, 于洁. 基于矢量跟踪的 SINS/GPS 深组合导航方法[J]. 中国惯性技术学报, 2009, 17(6): 710-717.
- [8] 谢非, 刘建业, 李荣冰, 等. 基于环路相关积分观测的 SINS/GPS 深组合导航算法[J]. 中国惯性技术学报, 2013, 21(4): 472-477.
- [9] 何晓峰. 北斗/微惯导组合导航方法研究[D]. 长沙: 国防科技大学, 2009.
- [10] GUSTAFSON D, DOWDLE J, FLUECKIGER K. A High Anti-Jam GPS based Navigator[C]. Proceedings of ION National Technical Meeting, Anaheim, USA, 2000.

(责任编辑 杨继森)

(上接第 62 页)

参考文献:

[1] 代东升, 贾迪阳, 谢峰. 面向装备保障的数据体系分析研究[J]. 四川兵工学报, 2015(6): 58-60.

[2] 王向博, 贾红丽, 刘钢, 等. 基于数据挖掘的复杂装备维修辅助决策研究[J]. 计算机与数字工程, 2012, 40(8): 142-145.

[3] 林平, 刘永辉, 陈大勇. 军事数据工程基本问题分析[J]. 军事运筹与系统工程, 2012, 26(1): 14-17.

[4] BENKHELIFA E, WELSH T, TAWALBEH L, et al. Creating evolving user behavior profiles automatically[J]. IEEE Transactions on Knowledge and Data Engineering, 2012, 24(5): 854-867.

[5] 余孟杰. 产品研发中用户画像的数据建模[J]. 设计艺术研究, 2014, 4(6): 60-64.

- [6] BENKHELIFA E, WELSH T, TAWALBEH L, et al. User profiling for energy optimisation in mobile cloud computing [J]. Procedia Computer Science, 2015, 52: 1159-1165.
- [7] DAM J W, VELDEN M. Online profiling and clustering of Facebook users[J]. Decision Support Systems, 2015, 70: 60-72.
- [8] 李林, 吴跃, 叶茂. 基于概率图模型的图像整体场景理解特征工程综述[J]. 计算机应用研究, 2015, 32(12): 3542-3550.
- [9] 张学工. 模式识别[M]. 北京: 清华大学出版社, 2010.
- [10] 申德荣, 于戈, 王习特, 等. 支持大数据管理的 NoSQL 系统研究综述[J]. 软件学报, 2013, 24(8): 1786-1803.
- [11] 朱明. 数据挖掘导论[M]. 合肥: 中国科学技术大学出版社, 2012.

(责任编辑 唐定国)