

doi: 10.3969/j.issn.1006-1576.2010.03.001

基于关联规则的武器装备体系能力分类

黄魏, 田亮, 杨克巍, 高兵

(国防科学技术大学 信息系统与管理学院, 湖南 长沙 410073)

摘要: 针对武器装备体系能力分类涉及利益相关者众多的特点, 提出基于关联规则的武器装备体系能力分类方法。介绍关联规则的基本概念, 对武器装备体系能力进行预处理后, 采用 FP growth 算法挖掘能力间的关联规则, 然后对产生的关联规则剪枝, 根据关联规则完成能力的分类, 通过编程对基于规则的武器装备体系能力分类进行测试。分类实验表明, 该方法可以完成能力的自动分类, 并具有良好的准确率和召回率。

关键词: 关联规则; FP 树; 能力; 分类

中图分类号: C934 **文献标识码:** A

Capability Categorization of Weapons Equipment System Based on Association Rules

HUANG Wei, TIAN Liang, YANG Ke-wei, GAO Bing

(School of Information System & Management, National University of Defense Technology, Changsha 410073, China)

Abstract: Weapons capability categorization system for the stakeholders involved in a large number of features, put forward a new method of capability categorization of weapons equipment system based on association rules. Introduce the basic concepts of association rules, after preprocessing the weapons equipment system capability, use the FP-Growth algorithm to mine the association rules, and filtrated the association rules. Then, based on complete ability categorization, test the proposed method by programming. The experimental results show that the method can achieve the automatic categorization of capabilities and has better accuracy and recalling rate.

Keywords: Association rules; FP-tree; Capability; Categorization

0 引言

武器装备体系能力是武器装备体系需求开发的核心要素^[1-2], 面向需求的能力是利益相关者对于武器装备体系完成特定战略规划的希望与约束, 表现为武器装备体系执行或完成特定使命任务所具备的“本领”或具有的“潜力”^[3]。传统的能力分类多是人工完成, 而武器装备体系分类结果存在不确定的问题, 不利于武器装备体系层面能力分类的达成共识, 而且人工分类也会耗费大量人力、物力和时间。故提出基于关联规则实现武器装备体系能力的自动分类, 以规范武器装备体系能力分类的过程。

1 关联规则

1.1 基本概念

关联是指两个或多个变量的取值或者活动之间存在某种规律性, 这种相互关联的规律可以理解为关联规则。关联规则用来描述在一个事务中交易之间同时出现的规律的知识模式, 这些规则展示属性—值频繁地在给定对象集中一起出现的条件, 关联规则挖掘可以发现大量数据中项集之间有趣的关

联或相关联系^[4]。

有 m 个不同项的集合 $I = \{i_1, i_2, \dots, i_m\}$, 其中的元素称为项, 记 D 为事务 T 的集合, 即 $D = \{T_1, T_2, \dots, T_n\}$, 事务 T 是项的集合, 并且 $T \subseteq I$ 。设 A, B 是一个 I 中项的集合, 如果 $A \subseteq T, B \subseteq T$, 那么称事务 T 包含 A, B 。一个关联规则是形如 $A \rightarrow B$ 的表达式, 其中 $A \subseteq I, B \subseteq I$ 且 $A \cap B = \emptyset$, 称 A 为规则的前项集 (前件), B 为规则的后项集 (后件)。

1.2 FP growth 算法^[5-6]

现有的关联规则挖掘算法多是以 Apriori 算法为基础的, 产生关联规则时需要生成大量的候选项目集。为了避免生成候选项目集, Han 等提出了基于 FP 树生成频繁项目集的 FP growth 算法, 该算法首先把频繁项目的挖掘问题转换成挖掘 FP 树问题, 然后使用 FP 树项目集增大挖掘频繁项目集。

FP 树挖掘过程可描述为: 由长度为 1 的频繁项目开始, 构造其条件项目基和条件 FP 树, 并递归地在该树上进行挖掘, 项目增长通过后缀项目与条件 FP 树产生的频繁项目连接实现。其中, FP 树

收稿日期: 2009-10-15; 修回日期: 2009-12-04

基金项目: “十一五”武器装备预先研究项目 (513300102); 国家自然科学基金 (70901074)

作者简介: 黄魏 (1982-), 男, 湖南人, 国防科技大学在读博士研究生, 从事复杂系统、体系需求建模技术、文本挖掘研究。

的构造过程可描述为：首先创建树的根结点，用“null”标记；扫描交易数据集 DB，每个事务中的项目按照支持度递减排序，并对每个事务创建一个分枝，通常当为一个事务考虑增加分枝时，沿共同前缀上的每个结点的计数值增加 1，为跟随在前缀之后的项目创建结点并链接；为方便遍历树，创建一个频繁项目列表，使得每个项目通过一个结点头指针指向它在树中的位置。

FP-growth 算法将发现大频繁项目集的问题转换成递归地发现一些小频繁项目，然后连接后缀。它使用最不频繁的项目后缀，提供了好的选择性，大大降低了搜索开销。

挖掘武器装备体系能力关联规则，可发现能力之间的关联关系，根据这种关联关系可判断多个能力之间存在一定的相似性，是能力分类的重要依据。

2 武器装备体系能力分类

在武器装备体系能力分析阶段，武器装备体系能力的外在表现形式是文本，文本内容即为利益相关者对武器装备体系能力的具体描述，包括其名称、承担的具体任务及其完成任务的手段方法等。因此，可以基于文本内容挖掘武器装备体系能力的关联规则，并以此为依据对武器装备体系能力进行分类。

基于关联规则的武器装备体系能力分类首先提取待测试武器装备体系能力文本的特征，而后根据用户指定的最小支持度发现频繁项集，根据最小置信度对产生关联规则并进行规则剪枝，最后根据修剪后的关联规则将测试能力分组到相应的类别中。

2.1 武器装备体系能力的预处理

武器装备体系能力预处理是指对从能力文本中抽取出的元数据即特征进行量化，以结构化形式描述文本信息，即将能力文本转化为特征词的集合。能力文本的特征词作为武器装备体系能力的中间表示形式，在分类时用以评价能力之间或能力与能力类之间的关联程度。

武器装备体系能力有类标签 C ，其特征集为 $T = \{t_1, t_2, \dots, t_m\}$ 。用 $D = \{C, t_1, t_2, \dots, t_m\}$ 表示该武器装备体系能力模型，则武器装备体系能力集可以表示成事务集，其项集由武器装备体系能力特征属性集（即词条）和武器装备体系能力所属类别组成，所有的关联规则都从事务集中产生。

2.2 基于 FP 树的关联规则挖掘

武器装备体系能力分类中感兴趣的关联规则

形如： $T \rightarrow C$ （其中 T 表示能力特征集， C 表示武器装备体系能力类标签）的事务。因此，首先要找到规则前件 T 中的频繁项集，即能力特征词，然后将频繁项集中的能力特征词与其类标签 C （规则后件）约束合并转为规则，最后计算规则集中每条规则的置信度，作为最后的结果保存。

基于 FP 树的武器装备体系能力关联规则挖掘算法实现的核心代码如下，其中的一些函数是系统自定义的。

```

this.Min_Sup = m_s*T_number();//计算最小支持数
Transaction newCapabilityT=new Transaction();
//生成新能力事务
this.T_Build(_listTraList, _listStrList1);
//构建训练文本事务库,生成1-项频繁集
this.Tidy(_listStrList1, _listTraList, _listTraList1);
//整理频集和事务库,生成频繁事务库
List<ItemTable> _list_TreeNode = new List<ItemTable>();
_list_TreeNode=this.CreatFP_Tree(_listStrList1, _listTraList);
//生成FP-Tree
List<List<List<FP_TreeNode>>> _list_list_list_Node = new
List<List<List<FP_TreeNode>>>();
//生成各节点到根节点的路径
foreach (ItemTable it in _list_TreeNode)
{
    List<List<FP_TreeNode>> _list_list_TreePath = new
List<List<FP_TreeNode>>();
    foreach (FP_TreeNode fp in it._List_FP_TreeNode)
        _list_list_TreePath.Add(fp.ConverToPath(fp));
    if (_list_list_TreePath.Count!=0)
        _list_list_list_Node.Add(_list_list_TreePath);
}
foreach (List<List<FP_TreeNode>> _list_list_TreePath1 in
_list_list_list_Node)
    this.Intersection(_list_list_TreePath1,(double)
this.Min_Sup);//生成同名节点的公共路径。
this.CreatRules(double min_conf);//生成关联规则

```

2.3 关联规则剪枝

关联规则 $T \rightarrow C$ 的作用度（lift）表示为 $l = P(C|T)/P(C)$ ^[7]：

- 1) 如果 $l(T \rightarrow C)$ 的值大于 1，称 T 和 C 正相关，表明每一个的出现都蕴含另一个的出现；
- 2) 如果 $l(T \rightarrow C)$ 的值等于 1，则 T 和 C 相互独立；
- 3) 如果 $l(T \rightarrow C)$ 的值小于 1，说明它们负相关。

因此，要保存的关联规则就是那些作用度大于 1 的关联规则。

对于通过作用度筛选后的规则，按照置信度、支持度优先级递减的方式从大到小排列，如果 2 个

规则的前件相同, 则保留置信度大的规则, 如果置信度相同, 则保留支持度大的规则^[8]。

2.4 能力分类

获得关联规则后, 就可以对体系能力进行分类。在提取待分类体系能力的特征集 T 后, 将生成的关联规则与 T 进行匹配, 此时会有匹配成功与失败 2 种情况。当匹配成功时分类完成, 当匹配失败时, 可以采用以下策略:

1) 抛弃关联规则, 按照最优匹配原则, 将待分类体系能力与训练能力文本中的体系能力进行匹配, 获得分类;

2) 以某步长为标准, 减小最小支持度(或置信度)的预定值, 再次进行分类;

3) 将待分类的体系能力添加到训练能力文本中, 对新训练能力文本进行聚类分析, 获得待分类的体系能力的类别。

基于关联规则的武器装备体系能力分类受训练文本、用于特征选择的统计量、规则选择原则等因素的影响, 它是一种相对分类, 而非绝对的。

3 实验结果

用美国海军能力领域中的能力文本的译本构建了分类的训练文本和测试文本, 通过编程对基于关于规则的武器装备体系能力分类方法进行了测试。分类程序界面如图 1。

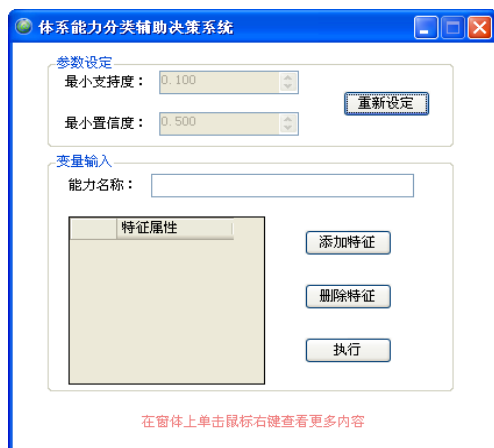


图 1 武器装备体系能力分类程序界面

输入水上作战及其特征属性后执行分类程序即可将其分组至海上防护能力类, 如图 2, 这个结果与美国海军能力领域中的分类是一致的。

采用所开发的程序对武器装备体系能力分类进行测试, 准确率为 86.7%, 召回率为 81.25%。

4 结束语

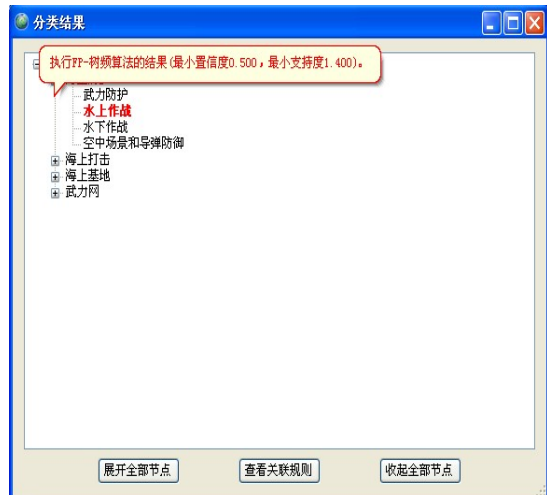


图 2 能力分类结果

基于关联规则的分类是一种经验分类, 对于军事领域有一定的风险性; 对于特殊问题, 需要发挥相关人员的主观能动性。该方法的主要作用在于形成武器装备体系能力分类的规范以获得利益相关者的广泛认可, 辅助决策。实验结果表明, 该方法基本满足能力分类的需要, 能发挥辅助决策作用。在下一步的研究中, 将研究能力与能力类之间的关联度, 进一步提高武器装备体系能力分类的准确率。

参考文献:

- [1] Joint Chief of staff. CJCSI3170.017D joint capabilities integration and development system[S/OL]. 2005-08-16. http://www.dtic.mil/cjcs_directives/index.htm
- [2] MODAF Project Review Board. MOD Architectural Framework Overview 1.0 [OB/OL]. 2005-8-31. <http://www.modaf.com>.
- [3] 赵青松, 黄巍, 鲁延京, 等. 基于概念格的体系使命任务与能力关系分析[J]. 系统仿真学报, 2009, 12(12): 3782-3784.
- [4] R. Agrawal, T. Imielinski, and A. Swami. Mining association rules between sets of items in large databases[R]. Proceedings of the ACM SIGMOD Conference on Management of data. 1993: 207-216.
- [5] 易彤, 徐宝文, 吴方君. 一种基于 FP 树的挖掘关联规则的增量更新算法[J]. 计算机学报, 2004, 27(5): 703-710.
- [6] Han J. et al. Mining frequent patterns without candidate generation[J]. In: Proceedings of the 2000 ACM SIGMOD Conference on Management of Data. Dallas. TX. 2000: 1-12.
- [7] 马光志, 张生庭. 基于关联规则的 Web 文档分类[J]. 计算机工程与设计, 2005, 26(9): 2515-2518.
- [8] 张帆, 严聪, 郭建亮. 改进灰色关联分析在导弹对抗预警雷达效能评估中的应用[J]. 四川兵工学报, 2009(9): 107-111.