面向特征的软件需求规约

张国生

(云南大学 信息学院，云南 昆明 650500)

摘　要：基于大数据方法对自然语言用户需求文本进行分析，用 TF-IDF 方法计算特征词在需求文本中的出现频率，归一化后求得特征词在文本中的权重，用卡方统计量衡量特征词和簇之间的相关性，用 k-均值算法和二分 k-均值算法对特征集聚类，生成用户需求文本特征树，并映射为需求功能分解树中对应的需求功能，使需求特征语义、结构更加清晰，增强了自动需求规约能力，提高了需求分析效率，使需求规约更加科学、合理、准确.

关键词：TF-IDF; 卡方; k-均值; 二分 k-均值; 特征树; 功能分解树

Feature-oriented software requirements specification

ZHANG Guo-sheng

(School of Information, Yunnan University, Kunming 650500,China)

Abstract：Based on big data method, the text of natural language user requirement is analyzed. TF-IDF method is used to calculate the frequency of feature words appearing in the requirement text. After normalization, the weight of feature words in the text is obtained. Chi-square statistics is used to measure the correlation between feature words and clusters. The feature set is clustered by k-means algorithm and bisecting k-means algorithm, and the user requirement text feature tree is generated, which is mapped to the corresponding requirement function in the requirement function decomposition tree. The semantics and structure of requirement feature are clearer. The ability of automatic requirement specification is enhanced. The efficiency of requirement analysis is improved, and the requirement specification is more scientific, reasonable and accurate.

Key words：TF-IDF; Chi-square; k-means; bisecting k-means; feature tree; function decomposition tree

作者简介：

张国生　男，（1968-），硕士，副教授.研究方向为软件工程.E-mail：zhanggs@ynu.edu.cn.