文章编号:1001 - 9081(2008)09 - 2353 - 04

基于灰色 Markov 模型动态关联规则的元规则挖掘

刘 俊,谢彦峰,张忠林,贾利敏
(兰州交通大学 电子与信息工程学院,兰州 730070)
(liujun_1024@163.com)

摘 要:介绍了增加了支持度向量和置信度向量两种规则评价指标的动态关联规则,给出了一种基于灰色 Markov 模型的预测和分析动态关联规则的元规则的方法。此方法在建立灰色模型的基础上应用 Markov 链理论,实验证明利用此方法挖掘的元规则要优于灰色模型等其他方法。

关键词:动态关联规则;元规则;灰色模型;Markov 连;预测
中图分类号: TP391 **文献标志码**:A

Research of mining meta-association rules for dynamic association rule based on model of Grey-Markov

LIU Jun, XIE Yan-feng, ZHANG Zhong-lin, JIA Li-min
(*School of Electronic and Information Engineering, Lanzhou Jiaotong University, Lanzhou Gansu* 730070, *China*)

Abstract: A rule called a dynamic association rule was introduced, which contained not only a support value and a confidence value but also a support vector and a confidence vector, and the usual mining process of the meta-association Rules for dynamic association rule by the model of Grey-Markov was introduced. This method applies the Markov theory in the foundation of G(1. 1) model, and proved that this method is superior to G(1. 1) and other methods in mining the meta-association rules.

Key words: dynamic association rule; meta-association rule; gray model; Markov chains; forecast

0 引言

关联规则是一个重要的数据挖掘研究课题[1],它主要用于发现事务数据集中项与项之间的关系,为决策提供参考。在关联规则挖掘中,已经产生的传统算法都是基于规则的支持度和置信度两个重要指标,并认为发现的关联规则在数据库中是永恒有效的,没有考虑到规则的变化,而由于事务数据通常具有时间特性,规则随着时间的推移可能会有很大的变化,因此考虑规则的变化更加符合规则的实际特性,为规则建立元规则可以更加直观地描述规则变化趋势,更好地为决策服务。

本文在文献[2]提出的增加了支持度向量和置信度向量两个新的规则评价指标的动态关联规则的基础上,利用灰色 Markov 模型对动态规则的元规则进行了分析和预测。

1 相关定义和定理

1.1 动态关联规则的定义

动态关联规则是一种能够描述自身特性随时间变化的关联规则[2-3],描述如下:

设 $I = \{i_1, i_2, \cdots, i_m\}$ 是项集合,任务相关的事务数据集 D 是在时间段 t 内收集到的,t 可分为不相交的长度为 n 的时间序列,即有 $t = \{t_1, t_2, \cdots, t_n\}$。同时根据 t 的划分,整个数据集 D 可以被分为 n 个数据子集:$D = \{D_1, D_2, \cdots, D_n\}$,其中数据子集 $D_i(i \in \{1,2,\cdots,n\})$ 的数据是在 $t_i(i \in \{1,2,\cdots, n\})$ 时间段内收集的。项集 T 满足 $T \subseteq I$。若 A 和 B 为项集,$A \subset I, B \subset I$,并且 $A \cap B = \varnothing$,则有如下动态关联规则相关定义。

*支持度向量(**SV**)* 动态关联规则 $A \Rightarrow B$(或者项集 $A \cup B$)的支持度向量,具有如下的表示形式:

$$\boldsymbol{SV} = [s_{(A \cup B)_1}, s_{(A \cup B)_2}, \cdots, s_{(A \cup B)_n}]$$

$$\text{s.t.} \quad s_{(A \cup B)_i} = \frac{f_{(A \cup B)_i}}{M}; i \in \{1,2,\cdots n\} \tag{1}$$

其中:$f_{(A \cup B)_i}$ 为项集 $A \cup B$ 在数据子集 $D_i(i \in \{1,2,\cdots,n\})$ 中出现的频数,M 为 D 中的事务数。设项集 $A \cup B$ 的支持度为 s 则有:

$$s = \sum_{i=1}^{n} s_{(A \cup B)_i} \tag{2}$$

有时,利用项集出现的频数表示支持度更为合适,这样项集的支持度向量为 $\boldsymbol{SV} = [f_1, f_2, \cdots, f_n]$,相应的支持度可以表示为:$s = \sum_{i=1}^{n} f_i$。

*置信度向量(**CV**)* 动态关联规则 $A \Rightarrow B$ 的置信度向量具有如下的表示形式:

$$\boldsymbol{CV} = [c_{(A \cup B)_1}, c_{(A \cup B)_2}, \cdots, c_{(A \cup B)_n}]$$

$$\text{s.t.} \quad c_{(A \cup B)_i} = \frac{s_{(A \cup B)_i}}{\sum_{i=1}^{n} s_{A_i}} = \frac{s_{(A \cup B)_i}}{s_A} \tag{3}$$

其中:$s_{(A \cup B)_i}$ 为项集 $A \cup B$ 的 $\boldsymbol{SV}$ 中的第 i 个元素,s_{A_i} 为项集

收稿日期:2008 - 04 - 08;**修回日期**:2008 - 05 - 10。
作者简介:刘俊(1985 -),男,河南固始人,硕士研究生,主要研究方向:数据挖掘; 谢彦峰(1951 -),男,河北石家庄,高级工程师,主要研究方向:网络管理、系统仿真、嵌入式相关技术; 张忠林(1965 -),男,河北阜城人,教授,博士,主要研究方向:软件工程、数据挖掘; 贾利敏(1981 -),女,河北邯郸人,硕士研究生,主要研究方向:软件工程。

A 的 $\boldsymbol{SV}$ 中的第 i 个元素，s_A 为项集 A 的支持度。

根据上述定义，关联规则 $A \Rightarrow B$ 的置信度 c 可以通过式(4)计算得到：

$$c = \frac{s_{(A \cup B)}}{s_A} = \frac{\sum_{i=1}^{n} s_{(A \cup B)_i}}{S_A} = \sum_{i=1}^{n} c_{(A \cup B)_i} \tag{4}$$

这样一条完整的动态关联规则就可以表述为：具有支持度向量 $\boldsymbol{SV}$，置信度向量 $\boldsymbol{CV}$，支持度 s，置信度 c 四个参数的关联规则。它具有如下表示形式：$A \Rightarrow B(\boldsymbol{SV}, \boldsymbol{CV}, s, c)$，其中 $\boldsymbol{SV}, \boldsymbol{CV}, s, c$ 可以分别根据式(1)～(4)计算，并一起来描述关联规则的动态性质。

1.2 元规则的定义

一般地，元规则[4]形成一个关于用户希望探察或证实的、他感兴趣联系的假定。然后，挖掘系统可以寻找与给定元规则匹配的规则。元规则形如：$P_1 \wedge P_2 \wedge \cdots \wedge P_l \Rightarrow Q_1 \wedge Q_2 \wedge \cdots \wedge Q_r$ 的规则模板。其中，$P_i(i = 1,2,\cdots,l)$ 和 $Q_j(j = 1,2,\cdots,r)$ 是例示谓词或谓词变量。

元规则是“规则的规则”，是对数据间依赖关系的关联规则的形式的模式，它使用户可以说明他们感兴趣规则的语法形式，规则的形式可以做为约束，指导知识的发现，帮助提高挖掘过程的性能。元规则可以根据分析者的经验、期望或对数据的直觉或者数据库模式自动生成。

1.3 灰色 G(1.1)预测模型

灰色 G(1.1)预测模型的原理是：对原始序列进行累加(或累减)，使生成的序列呈一定规律，其对应的曲线(或折线)可以用典型曲线逼近，然后用逼近的曲线作为模型，用以对系统进行预测。其具体步骤如下。

1)设原始序列为 $\{x^{(0)}(i)\}(i = 1,2,\cdots,n)$，对原始序列进行累加生成序列：

$$x^{(1)}(i) = \sum_{j=1}^{i} x^{(0)}(j) \tag{5}$$

2)设 $\{x^{(1)}(i)\}(i = 1,2,\cdots,n)$ 满足单变量常微分方程：$\frac{\mathrm{d}x^{(1)}}{\mathrm{d}t} + ax^{(1)} = u$，其中 a 为发展数，u 为灰色作用变量，此方程称为 $GM(1,1)$ 模型的白化方程，它的解为 $x^{(1)}(t) = (x^{(1)}(t_0) - \frac{u}{a})e^{-a(t-t_0)} + \frac{u}{a}$，时间间隔取离散值为：

$$x^{(1)}(k+1) = (x^{(1)}(1) - \frac{u}{a})e^{-ak} + \frac{u}{a} \tag{6}$$

3)通过最小二乘法来估计上面方程中 a 和 u，令：

$$\overline{Y} = \begin{bmatrix} x^{(0)}(2) \\ x^{(0)}(3) \\ \vdots \\ x^{(0)}(n) \end{bmatrix} \tag{7}$$

$$B = \begin{bmatrix} -\frac{1}{2}(x^{(1)}(2) + x^{(1)}(1)) & 1 \\ -\frac{1}{2}(x^{(1)}(3) + x^{(1)}(2)) & 1 \\ \vdots & \vdots \\ -\frac{1}{2}(x^{(1)}(n) + x^{(1)}(n-1)) & 1 \end{bmatrix} \tag{8}$$

则最小二乘法的估计值为：

$$\hat{\overline{Y}} = \begin{bmatrix} \hat{a} \\ \hat{u} \end{bmatrix} = (\boldsymbol{B}^{\mathrm{T}}\boldsymbol{B})^{-1}\boldsymbol{B}^{\mathrm{T}}\,\overline{Y} \tag{9}$$

4)把估计值 $\hat{a}, \hat{u}$ 代入式(6)中得到时间影响方程：

$$\hat{x}^{(1)}(k+1) = (x^{(1)}(1) - \frac{\hat{u}}{\hat{a}})e^{-\hat{a}k} + \frac{\hat{u}}{\hat{a}} \tag{10}$$

当 $k = 1,2,\cdots,n-1$ 时，由式(8)得到的是拟合值，而当 $k \geqslant n$ 时，得到的是预测值。

5)对 $\hat{x}^{(1)}(k+1)$ 做累减还原处理可得原始数据预测公式：

$$\hat{x}^{(0)}(k+1) = \hat{x}^{(1)}(k+1) - \hat{x}^{(1)}(1) \tag{11}$$

1.4 Markov 链及其转移概率矩阵

Markov 链[5] 设有随机过程 $\{X_n \mid n \in T\}$，对任意的整数 $n \in T$ 和任意的 $i_0, i_1, \cdots, i_{n+1} \in I$，有 $P\{X_{n+1} = i_{n+1} \mid X_0 = i_0, X_1 = i_1, \cdots, X_n = i_n\} = P\{X_{n+1} = i_{n+1} \mid X_n = i_n\}$，则称 $\{X_n \mid n \in T\}$ 为 Markov 链，简称马氏链。其中，参数集 T 是离散的时间集合，即 $T = \{0,1,2,\cdots\}$，其相应的 X_n 可能取值的全体组成的状态空间是离散的状态集 $I = \{i_1, i_2, i_3, \cdots\}$。

转移概率 称条件概率 $p_{ij}(n) = P\{X_{n+1} = j \mid X_n = i\}$ 为 Markov 链 $\{X_n, n \in T\}$ 在时刻 n 的一步转移概率，其中 $i, j \in I$，简称为转移概率。

一步转移概率矩阵 如果 Markov 链的状态空间 $I = \{1, 2, \cdots\}$，则以一步转移概率 p_{ij} 所组成的矩阵 $\boldsymbol{P}$ 称为系统状态的一步转移概率矩阵。

$$\boldsymbol{P} = \boldsymbol{P}(1) = \begin{bmatrix} p_{11} & p_{12} & \cdots \\ p_{21} & p_{22} & \cdots \\ \vdots & \vdots & \end{bmatrix} \tag{12}$$

显然，矩阵 $\boldsymbol{P}$ 的所有元素都是非负的，且每一行元素的和都等于1。

且普曼-柯尔莫哥洛夫方程(C-K 方程) 设 $\{X(n) \mid n \in T_1\}$ 是一齐次马氏链，则对任意的 $u, v \in T_1$，有 $P_{ij}(u+v) = \sum_{k=1}^{\infty} P_{ik}(u)p_{kj}(v), i, j = 1,2,\cdots$。

C-K 方程的矩阵形式 $P(u+v) = P(u)P(v)$。利用C-K方程确定 n 步转移概率，在 $P(u+v) = P(u)P(v)$ 中令 $u = 1, v = n-1$，得递推关系：$P(n) = P(1)P(n-1) = PP(n-1)$，从而可得：

$$P(n) = P^n \tag{13}$$

这说明马氏链的 n 步转移概率是一步转移概率的 n 次方，链的有限维分布可由初始分布和一步转移概率完全确定。

2 动态关联规则元规则挖掘描述

现有的关联规则挖掘算法大多是基于给定的最小支持度和最小置信度的静态关联规则挖掘算法，这样做的前提假设是：数据集中各项具有近似的性质和作用，即重要性相同或相近；数据集中各项的分布是均匀的，即出现的频率相近或相似[6]。然而，在现实的数据库中却往往并非如此，对用户来说，不同项的价值和重要程度是不一样的，不同的项在数据集中出现的频繁程度也不一样，而且事实上不频繁的项并不一定不重要，因此利用统一的最小支持度进行整个数据集的不同时段的关联规则挖掘是不充分的。由1.1节中动态关联规则的定义可知，根据不同时间段划分出不同的子数据集，并在

此基础上定义了支持度向量和置信度向量,通过支持度向量、置信度向量、支持度和置信度四个参数共同描述规则的动态性质,能够提供规则与时间相关的信息,这样便能在一定程度上解决上诉不足之处。

但动态关联规则仍然存在如何选取支持度的问题,而元规则挖掘的目的之一是发现关联规则序列随时间变化的趋势,通过对趋势的分析,预测下一个时间段的支持度和置信度的可能值,因此为动态关联规则建立了元规则,在元规则的指导下,可以对数据集中的关联模式进行更加准确有效的挖掘。目前提出建立元规则的方法主要有基于概率统计的方法[7]和基于模糊决策树的方法[8]。基于概率的方法主要是采用回归分析等对规则的支持度进行曲线拟合,这在处理不确定数据上效果欠佳;而从批量处理规则的角度出发,由于需要较多的专家信息,基于模糊决策树的方法明显无法满足要求。本文提出一种基于灰色 Markov 模型的方法,可以很好的解决这些问题,具体步骤如下:

1)给定的支持度向量为 $\boldsymbol{SV} = [f_1,f_2,\cdots,f_n]$,其中 f_i 为规则在时间段 t_i 内出现的频数。将序列 $\{f_1,f_2,\cdots,f_n\}$ 看作原始序列 $\{x^{(0)}(i)\}$ $(\mathrm{i}=1,2,\cdots,\mathrm{n})$,建立 GM(1.1)模型。

2)利用 1.3 节中介绍的方法得到原始数据预测公式(11)。

3)利用式(11)预测得到序列 $\{x^{(0)}(i)\}$ $(\mathrm{i}=1,2,\cdots,\mathrm{n})$ 的拟合值 $\{\hat{x}^{(0)}(i)\}$ $(\mathrm{i}=1,2,\cdots,\mathrm{n})$,并通过计算 $\{\hat{x}^{(0)}(i)\}-x^{(0)}(i)$ 得到拟合值与实际值的差额,以及差额所占的百分比。

4)根据各个支持度频数的预测结果和此支持度向量序列的增幅及 Markov 链分析方法,结合实际情况对差额所占百分比的情况进行状态划分。

5)采用频率接近概率的方法计算每一个状态转移到其他任何状态的转移概率 P_{ij},进而得到一步转移概率矩阵 $\boldsymbol{P}$ 和 k 步转移概率矩阵 $\boldsymbol{P}(k)$。

6)通过一步转移概率矩阵预测下一状态,及通过 k 步转移概率矩阵预测当前状态下一状态开始的第 k 个状态,通过预测便可得知此支持度向量下一值或下一值开始的第 k 值取值的概率。

7)确定预测对象未来的状态转移后,即确定了预测值的变动灰色区间,取灰区间的中值作为最终的预测值。

3　实例分析

为了直观的说明上诉方法在动态关联规则的元规则挖掘中预测支持度值的具体过程,下面将针对一实例用基于灰色 Markov 模型方法对其支持度序列的值进行预测。

实例　设规则 $A \Rightarrow B$ 表示,顾客在购买 A 的前提下购买 B,且它在 1998 ~ 2007 年这十年内的支持度计数(即:事件顾客购同时购买了 A 和 B 发生的次数)f_i 构成了该规则的支持度向量:$\boldsymbol{SV}$ = [79,90,99,117,154,201, 248,335,402, 498],下面将采用灰色 *Markov* 模型对该规则在 2008 ~ 2010 年这三年的支持度计数进行预测。

为了验证此模型的预测准确性,用此模型预测 2007 ~ 2010 年的支持度计数的预测值,将 2007 年的预测值与实际值相比较后,可以说明所预测的 2008 ~ 2010 年的数据是可靠的。

按以下步骤进行:

1)将给定序列{79,90,99,117,154,201, 248,335,402}看作原始序列 $\{x^{(0)}(i)\}(i=1,2,\cdots,9)$,建立 GM(1.1)模型。

2)利用 1.3 节中介绍的方法得到 a,u 的预测值 $\hat{a} = -0.229235$,$\hat{u} = 52.5925$,从而得到时间影响方程 $\hat{x}^{(1)}(k+1) = 308.426135\mathrm{e}^{0.229235k} - 229.426135$。

3)利用式(11)得到表 1,该规则在 1998 ~ 2006 年间的支持度计数的实际值,预测值(由于是支持度频数,故取整数),以及差额和差额所占百分比。

表 1　1998 ~ 2006 年 GM(1.1)预测结果对照

年份	实际值	预测值	差额	百分比/%
1998	79	79	0	0.0000
1999	90	84	-6	-6.6670
2000	99	100	1	1.0101
2001	117	125	8	6.8370
2002	154	158	4	2.5970
2003	201	199	-2	-0.9950
2004	248	250	2	0.8065
2005	335	314	-1	-1.2990
2006	402	393	-9	-2.2390

通过计算表中数据可知:根据这个灰色预测曲线方程得到的 1998 ~ 2006 年的预测值的平均误差为 1.4555%,而年度最大误差为:6.837%,故此灰色模型预测的结果可以接受。

4)根据该规则的支持度频数拟合结果和此支持度序列的增幅以及 Markov 链分析方法,依据差额百分比划分出 E_1, E_2, E_3, E_4, E_5 五种状态。

状态 E_1:百分比小于 -8%,由表 1 可知,在 1999 ~ 2006 年(1998 年数据非预测得到)内这种状态没有出现过;

状态 E_2:百分比大于 -8% 而小于 -2%,由表 1 可知,在 1999 ~ 2006 年内有 1999 和 2006 年呈现此状态;

状态 E_3:百分比大于 -2% 而小于 2%,由表 1 可知,在 1999 ~ 2006 年内有 2000、2003、2004、2005 呈现此状态;

状态 E_4:百分比大于 2% 而小于 8%,由表 1 可知,在 1999 ~ 2006 年内 2001 和 2002 年两年呈现此状态。

状态 E_5:百分比大于 8%,由表 1 可知,在 1999 ~ 2006 年内这种状态没有出现过。

5)计算状态转移概率,确定状态转移矩阵。由上分析可知,此规则在 1999 ~ 2006 十三年中各状态的转移情况如表 2 所示。

表 2　1999 ~ 2006 年内各状态转移情况

状态	状态 E_2	状态 E_3	状态 E_4	合计
E_2	0	1	0	1
E_3	1	2	1	4
E_4	0	1	1	2

由表 2 可得一步转移概率矩阵:

$$\boldsymbol{P} = \begin{bmatrix} 0 & 1 & 0 \\ 0.25 & 0.50 & 0.25 \\ 0 & 0.50 & 0.50 \end{bmatrix} \tag{14}$$

由式(13)可得:

$$P(2) = \begin{bmatrix} 0.250 & 0.500 & 0.250 \\ 0.125 & 0.625 & 0.250 \\ 0.125 & 0.500 & 0.375 \end{bmatrix} \tag{15}$$

$$P(3) = \begin{bmatrix} 0.12500 & 0.62500 & 0.25000 \\ 0.15625 & 0.56250 & 0.28125 \\ 0.12500 & 0.56250 & 0.31250 \end{bmatrix} \tag{16}$$

$$P(4) = \begin{bmatrix} 0.15625 & 0.56250 & 0.28125 \\ 0.14063 & 0.57812 & 0.28125 \\ 0.14063 & 0.56250 & 0.29687 \end{bmatrix} \tag{17}$$

6)由上各转移概率矩阵可得该规则在 2007 ~ 2010 年的支持度计数的各可能状态的概率如表 3 所示。

表 3 2007 ~ 2010 年此规则支持度计数可能状态的概率取值

状态	2007 年	2008 年	2009 年	2010 年
E_2	0.25000	0.12500	0.15625	0.14063
E_3	0.50000	0.62500	0.56250	0.57812
E_4	0.25000	0.25000	0.28125	0.28125

(7)取灰色区间的中值作为最终的预测值,2007 ~ 2010 年此规则支持度计数的预测结果如表 4 所示。

表 4 利用灰色 Markov 模型预测的结果

年份	灰色区间	预测中值	状态概率
2007	[488, 508]	498	0.50000
2008	[612, 638]	615	0.62500
2009	[771, 802]	786	0.56250
2010	[969, 1009]	989	0.57812

从表 4 可知,利用此模型得到的 2007 年的灰色区间为[488,508],预测中值为 498,而直接用灰色模型(时间影响方程)得到的值为 497,而实际值为 498,这说明此模型得到的值比直接用灰色模型预测值要可靠,故所得 2008 ~ 2010 年的数据是准确的。另外,在历史数据较多时,可根据实际情况划分出更详细的状态,这样可以使预测结果更加准确。从表 4 还可以看出此规则的支持度未来三年呈上升趋势。利用灰色 Markov 模型不但可以预测年份的支持度区间,还可以了解这样区间产生的概率,由预测中值和状态概率可以准确的把握动态规则支持度的未来趋势。

4 结语

本文在文献[2]提出的动态关联规则基础上,将动态关联规则与元规则挖掘相结合,提出了一种基于灰色 Markov 模型的动态关联规则的元规则挖掘方法,为动态关联规则建立元规则,利用此方法解决了动态关联规则过程中如何选取支持度的问题,克服了静态关联规则挖掘算法的数据集中各项差异较大或分布不均匀的缺点。通过一个实例说明利用本模型挖掘动态关联规则的元规则的一般过程,并通过预测值与实际值的比较得出利用此模型得到的值比直接用灰色模型预测值要可靠。利用本模型进行元规则挖掘可以得到规则序列未来变化的几种不同可能,比单纯使用灰色 G(1.1)模型预测一个值的结果要更加合理,且当数据的指数分布不明显时不能直接使用灰色 G(1.1)模型,但本模型由于在灰色模型的基础上使用了 Markov 知识,比灰色 G(1.1)模型的应用范围要广,总体效果要优于灰色 G(1.1)模型,能更准确的把握规则的变化趋势从而使动态关联规则挖掘在合理的元规则指导下得到更精确的结果。

参考文献:

[1] AGRAWAL R , SRIKANT R. Fast algorithms for mining association rules in large databases[C] // Proceedings of the 20th International Conference on VLDB. San Francisco: Morgan Kaufmann Publishers Inc., 1994: 487 - 499.

[2] 荣冈, 刘进锋, 顾海杰. 数据库中动态关联规则的挖掘[J]. 控制理论与应用, 2007, 24(1): 127 - 131.

[3] 沈斌, 姚敏. 一种新的动态关联规则及其挖掘算法研究[J]. 中国科技论文在线, 2007.

[4] HAN J W, KAMBER M. 数据挖掘: 概念与技术[M]. 范明, 孟晓峰, 译. 北京: 机械工业出版社, 2007: 147 - 183.

[5] 刘次华. 随机过程及其应用[M]. 北京: 高等教育出版社, 2004: 48 - 86.

[6] 欧阳为民, 郑诚, 蔡庆生. 数据库中加权关联规则的发现[J]. 软件学报, 2001, 12(4): 612 - 619.

[7] LIU BING, MA YI-MING, LEE R. Analyzing the interestingness of association rules from the temporal dimension [C]//IEEE International Conference on Data Mining (ICDM 2001). Washington, DC: IEEE Computer Society, 2001: 377 - 384.

[8] AU W-H, CHAN K C C. Mining changes in association rules: A fuzzy approach [J]. Fuzzy sets and systems, 2005, 149: 87 - 104.

(上接第 2352 页)

[7] LEE K, KANG K C. Feature dependency analysis for product line [C]// Proceedings of 8th International Conference, ICSR 2004. Madrid, Spain: [s. n.], 2004: 69 - 85.

[8] JARING M, BOSCH J. A taxonomy and hierarchy of variability dependencies in software product family engineering[C]// Proceedings of the 28th Annual International Computer Software and Applications Conference (COMPSAC' 04). Washington, DC: IEEE Computer Society, 2004: 356 - 361.

[9] LEE K, KANG K C, LEE J. Concepts and guidelines of feature modeling for product line software engineering[C]// Proceedings of software reuse: methods, techniques, and tools. Berlin: Springer, 2002, 2319: 62 - 77.

[10] GRISS M L, FAVARO J, D' ALESSANDRO M. Integrating feature modeling with the RSEB[C]// Proceedings of the 5th International Conference on Software Reuse. Washington, DC: IEEE Computer Society, 1998: 76 - 85.

[11] JACOBSON I, GRISS M, JONSSON P. Software reuse — architecture, process, and organization for business success[M]. [S. l.]: Addison-Wesley, 1997.

[12] CZARNECKI K, EISENECKER U W. Generative Programming: Methods, Tools, and Applications[M]. [S. l.]: Addison-Wesley. 2000.

[13] SCHOBBENS P-Y, HEYMANS P, TRIGAUX J-C. Generic semantics of feature diagrams[J]. Computer Networks, 2007, 51: 456 - 479.