

基于 Apriori 算法的浙西杉木用材林立地及生长因子关联分析

董晨^{1,2}, 夏凯^{1,2}

(1. 浙江农林大学 信息工程学院, 浙江 杭州 311300; 2. 浙江农林大学 浙江省林业智能监测与信息技术研究重点实验室, 浙江 杭州 311300)

摘要: 森林立地条件影响着林木的生长。为了充分挖掘立地因子与林分生长之间的内在联系, 以杭州市临安区森林资源动态监测数据为基础, 以杉木 *Cunninghamia lanceolata* 林为研究对象, 运用差分方程法构建了基于 Richard 理论方程的地位指数模型; 利用 Apriori 算法进行立地因子和林分生长因子的关联规则挖掘, 得到 175 条置信度在 80% 以上、支持度在 10% 以上的因子关联规则, 并从中提取了 20 条“林分因子-立地因子”规则。结果表明: 影响临安区杉木用材林立地质量的主要因子是海拔、坡位、坡向、坡度级、土壤类型、土壤质地、土层厚度、腐殖质层厚度、林下植被种类和林下植被覆盖度, 各类规则揭示了杉木林分立地因子和立地质量、立地因子和林分生长因子, 林分因子和林龄之间存在的变化规律以及隐含的关联关系。表 3 参 31

关键词: 森林经理学; Apriori 算法; 关联规则; 立地因子; 生长因子; 地位指数模型; 杭州市临安区
中图分类号: S758.5 **文献标志码:** A **文章编号:** 2095-0756(2019)04-0741-08

Chinese fir forest site association analysis and growth factors in western Zhejiang based on the Apriori algorithm

DONG Chen^{1,2}, XIA Kai^{1,2}

(1. College of Information Engineering, Zhejiang A&F University, Hangzhou 311300, Zhejiang, China; 2. Zhejiang Provincial Key Laboratory of Forestry Intelligent Monitoring and Information Technology, Zhejiang A&F University, Hangzhou 311300, Zhejiang, China)

Abstract: Forest site conditions affect the growth of trees. To fully delve into the internal relationship between site factors and stand growth factors, Chinese fir (*Cunninghamia lanceolata*) forest was taken as the research object, and a site index model based on Richard's theoretical equation was established using dynamic monitoring data of forest resources in Lin'an District of Hangzhou. The Apriori algorithm was used to explore the association rules between site and forest stand growth factors. Results showed 175 rules obtained with confidence over 80% and support over 10% and 20 rules of “stand growth factors-site factors” were extracted from them. The site factors which influenced site quality of Chinese fir timber forests in Lin'an District were altitude, slope position, slope aspect, gradient, soil type, soil texture, soil thickness, humus depth, species in the undergrowth, and degree of cover on the undergrowth through the rules. The rules revealed variation rules and hidden associations between site factors and stand quality, site factors and stand growth factors, as well as stand factors and stand age. [Ch, 3 tab. 31 ref.]

Key words: forest management; Apriori algorithm; association rules; site factors; stand growth factors; site in-

收稿日期: 2018-07-30; 修回日期: 2018-11-10

基金项目: 浙江省教育厅资助项目(Y201636009); 浙江农林大学科研发展基金项目(2016FR036); 浙江省科技重点研发计划资助项目(2018C02013); 浙江省自然科学基金委员会-青山湖科技城管委会联合基金项目(LQY18C160002)

作者简介: 董晨, 讲师, 博士, 从事森林资源管理与决策研究。E-mail: dongchen@zafu.edu.cn

dex model; Lin'an District in Hangzhou

在森林经营中,立地质量是衡量某一立地上森林生长环境以及植被生产潜力的重要指标,对林分的生长收获具有重要影响。不同立地条件的林分,林木的生长状态不尽相同,从而影响林分的最终收获。构成立地条件的因素很多,坡度、坡向、海拔等是影响林木生长的重要因子,选择正确的立地因子,并能找出主要立地因子对林分生长影响的规律性,这对于林木的引种栽培、适地适树、造林地选择以及科学评价林分的立地质量都有重要的实践意义。目前,关于立地因子与林木生长相关研究的报道已有不少,研究方法主要集中为2类:一是根据不同的立地因子对林分调查数据进行分组分类,单独分析每项立地因子不同分类下林木生长的差异^[1-3];二是通过构建数量化模型来探索立地因子与林木林分生长因子之间的相关性,在这方面的研究中,多见于构建林木生长因子和立地因子的多元回归模型来体现^[4-5]。然而,目前已有的立地因子与林木生长因子的相关性研究仅简单分析了部分立地因子与生长因子的统计学相关程度,并无量化其重要性程度及深入挖掘立地因子间存在的关联,同时,影响林木生长的立地因子也通常依靠经验进行选取,忽略了一些潜在的相关因子,使得研究结果具有较大的主观性。因此,寻找一种能快速自动地对立地因子与林分生长进行深入分析,从中挖掘出因子之间所隐藏关系的方法,对后期评估森林资源数据、指导林业发展显得非常必要。数据挖掘作为一个新兴的多学科交叉应用领域,是从大量、不完全、有噪声、模糊、随机的数据中,提取隐含在其中的、人们事先不知道但又是潜在有用的信息和知识的过程^[6-7],在森林资源管理与决策领域,数据挖掘技术在林分收获^[8]、林业遥感^[9]、生物量预测^[10]、林火模拟^[11-12]等方面有着一定的应用。数据挖掘中的一个重要任务是关联分析。关联规则挖掘是发现大量数据中项集之间一定的关联或相关联系^[13]。关联规则挖掘算法有数种,其中,Apriori算法由于其简单、易理解、数据要求低的特点,被广泛应用于各个领域^[14-15]。杉木 *Cunninghamia lanceolata* 在浙江省用材林中占有重要的地位^[16]。本研究以浙西杉木人工用材林为研究对象,使用 Apriori 算法对杉木用材立地因子及其生长指标进行关联分析,以期挖掘出立地质量与立地因子之间、以及立地因子与林分生长因子的关联规则,为当地立地质量评价以及立地模型的构建提供科学依据。

1 研究区概况与数据来源

1.1 研究区概况

临安区位于浙江省杭州市西部,29°56'~30°23'N,118°51'~119°52'E,地属中亚热带季风气候区南缘,属季风型气候,雨量充沛,年均降水量为1 613.9 mm,光照充足,立体气候明显,海拔从城市至天目山顶,年平均气温由16.0℃降至9.0℃,横跨亚热带和温带2个气候带。临安区地势自西北向东南倾斜,区境北、西、南三面环山,形成一个东南向的马蹄形屏障。西北多崇山峻岭,深沟幽谷;东南为丘陵宽谷,地势平坦,全境地貌以中低山丘陵为主。临安区森林覆盖率达78.2%,位居杭州各区县市之首;其森林面积达2 452 km²,是国家森林城市之一。临安区主要用材树种有杉木,马尾松 *Pinus massoniana*,湿地松 *Pinus elliottii*,水杉 *Metasequoia glyptostroboides*,响叶杨 *Populus adenopoda*,樟树 *Cinnamomum camphora*,柏木 *Cupressus funebris*,柳杉 *Cryptomeria fortunei*,檫木 *Sassafras tzumu*,枫香 *Liquidambar formosana*,毛竹 *Phyllostachys edulis* 等。

1.2 数据来源

源于临安区2008-2012年森林资源动态监测数据,在浙江省森林资源规划设计调查的基础上,建立县级森林资源年度动态监测体系。调查因子以小班为单位,包括了小班基本信息、立地因子、林分因子、权属信息、经营措施、病虫害火灾信息等。其中立地因子包括地貌、海拔、坡向、坡位、坡度级、土壤名称、土壤质地、土层厚度、腐殖质层厚度、林下植被种类、林下植被高度、林下植被覆盖度等12个。林分因子包括了林种、起源、树种、年龄、平均胸径、平均高、优势高、郁闭度、疏密度、单位株树、单位蓄积等。

在各类树种中,研究选取了2008和2012年每年的杉木人工纯林小班复测数据,其中地位指数模型构建选用2期的复测数据,林分生长与立地因子关联规则的研究则选用2012年的数据。由于早期杉苗处于恢复和扎根阶段,真正进入林木速生生长则是在5 a以后,同时,相关文献表明郁闭度在0.3以上

的林分才能充分体现林木的生长状态^[17], 因此本研究剔除年龄小于 5 a, 郁闭度小于 0.3 的小班数据。对数据的完整性及一致性进行检查, 以 3 倍标准差为标准剔除各项异常数据。经过数据整理, 得到 721 条供研究的小班数据, 分别分布于临安区玲珑、锦西等 54 个村。地貌为低山丘陵, 海拔 10~570 m。根据调查数据, 整理得到小班林分生长情况, 如表 1 所示。

2 研究方法

除了现有的立地因子之外, 还需要一个综合性衡量立地质量好坏的指标。评价某一有林地的立地质量时, 地位级、地位级指数和地位指数是常用的 3 种评定指标^[18]。地位指数具有特定的数学表达式, 且在建模过程中, 数据受人为干扰因素较少, 因而受到广泛地应用^[19-20]。因此, 本研究选择地位指数作为立地条件的判定指标。根据杉木用材林优势木高和年龄复测数据, 使用代数差分法构建杉木地位指数模型。根据模型计算出每个小班的地位指数, 将地位指数作为立地因子, 和其他 12 个立地因子一起, 联合林分生长因子胸径、树高和蓄积, 使用 Apriori 算法作关联规则分析。

2.1 地位指数模型构建

代数差分法(algebraic difference approach, ADA)是构建林分地位指数的常用方法之一, 其原理是选择一个理论方程作为基础方程, 选择方程中的一个参数作为消元参数, 对方程进行差分消元, 从而获得一个包含 2 组因变量和自变量的差分方程^[21]。研究选择 Richards 理论方程作为基础模型来构建差分方程, Richards 理论方程如式(1)所示:

$$H_t = a[1 - \exp(-ct)]^b \quad (1)$$

式(1)中: H_t 代表林分优势木高; t 为林分年龄; a, b, c 为林分参数, 其中 a 代表林木的潜在生长最大值, c 代表林木生长速率, b 代表消元参数。对其进行代数差分转换, 得到转换后的差分方程, 如式(2)所示:

$$H_{t_2} = a(H_{t_1}/a)^{\frac{\ln[1 - \exp(-ct_2)]}{\ln[1 - \exp(-ct_1)]}} \quad (2)$$

式(2)中: H_{t_1} 和 H_{t_2} 分别为 t_1 和 t_2 时刻的林分优势木高。最后根据小班的复测数据, 拟合得到地位指数模型为:

$$I_s = 12.941(H_t/12.941)^{\frac{\ln[1 - \exp(-0.141 \times 20)]}{\ln[1 - \exp(-0.141t)]}}, R^2 = 0.716 \quad (3)$$

式(3)中: I_s 为林分地位指数; H_t 为林分优势木平均高; t 为林分年龄。

2.2 数据预处理

在数据挖掘之前, 对原始数据进行数据清洗和数据转换。在本研究中, 1.2 节中的剔除异常数据和无效数据即数据清洗, 对清洗后的不同量纲数据进行归一化处理即数据转换。本研究数据包括定性因子和定量因子, 定性因子包含多个属性维度, 定量因子则为连续性数据。对定性因子按照类别进行分类, 对定量连续性数据进行离散化处理。数据具体转换过程如下: 地貌分为低山(A_1)、丘陵(A_2); 海拔分为 10~200 m(B_1), 201~400 m(B_2), 401~570 m(B_3)3 组; 坡位上、中、下、谷、全分别用 $C_1 \sim C_5$ 标识; 坡向东、南、西、北、东北、东南、西北、西南分别用 $D_1 \sim D_8$ 标识; 坡度级平、缓、斜、陡、急、险分别用 $E_1 \sim E_6$ 标识; 土壤分为红壤(F_1)、黄壤(F_2); 土壤质地分为砂土(G_1)、壤土(G_2)、黏土(G_3); 土层厚度分为厚(H_1)、中(H_2)、薄(H_3); 腐殖质厚度分为厚(I_1)、中(I_2)、薄(I_3); 林下植被种类分为草丛(J_1)、草灌(J_2)、灌丛(J_3)、无植被(J_4); 林下植被高度分为 0~50 cm(K_1), 51~85 cm(K_2); 林下植被覆盖度分为 0~30%(L_1), 31%~60%(L_2), 61%~90%(L_3)。选用最具代表性的胸径、树高和蓄积 3 个林分生长因子作为关联规则事务项, 作离散化处理。根据收集的数据范围, 将胸径分为 5.0~10.0 cm(M_1), 10.1~15.0 cm(M_2), 15.1~20.0 cm(M_3); 将树高分分为 3.0~6.0 m(N_1), 6.1~9.0 m(N_2), 9.1~12.0 m(N_3), 12.1~15.0 m(N_4), 15.1~18.0 m(N_5); 将单位蓄积分为 20.0~50.0 m³(O_1), 50.1~80.0 m³(O_2), 80.1~110.0 m³(O_3), 110.1~140.0 m³(O_4), 140.1~170.0 m³(O_5)。将林分年龄也考虑其中, 根据浙江省森林资源规划设计调查

表 1 浙西杉木林分生长基本概况

Table 1 General growth information of Chinese fir stands

项目	年龄/a	胸径/cm	树高/m	蓄积/(m ³ ·hm ⁻²)
最小值	5.0	5.9	3.6	22.0
最大值	38.0	18.8	17.7	164.0
平均值	21.4	15.6	9.8	88.0
标准差	9.38	4.87	4.09	49.44

规程,对浙西杉木用材林进行龄组划分, ≤ 10 a(幼龄林, P_1), 11~20 a(中龄林, P_2), 21~25 a(近成熟林, P_3)和 26~35 a(成熟林, P_4)和 > 35 a(过熟林, P_5)。根据构建的地位指数模型,计算得到杉木地位指数为 8~18,作为关联分析事项,将 16 和 18 指数的立地定为好,用 Q_1 表示;12 和 14 指数立地为中,用 Q_2 表示;8 和 10 指数立地为差,用 Q_3 表示。表 2 是依据分类进行预处理的部分数据,其中每行数据可以理解成 Apriori 算法中的一个事务。本研究首先使用 Apriori 算法计算出频繁项集及支持度,再根据置信度计算公式得到项集与项集之间的规则。

表 2 预处理后的林分数据

Table 2 Stand data after pretreatment

小班号	年龄	胸径	树高	地貌	海拔	坡位
004	P_3	M_2	N_2	A_2	B_1	C_5
011	P_2	M_2	N_1	A_1	B_1	C_1
018	P_3	M_3	N_3	A_1	B_2	C_1
022	P_1	M_1	N_1	A_1	B_1	C_1
.....

2.3 关联规则与 Apriori 算法

关联规则是对事物间或关系数据集中项之间的关联或相关性的描述,关联规则挖掘兴趣度的 2 种度量主要是支持度和置信度。计算方法分别如下:

$$\text{sup}R_1 = \frac{\text{count}(X \cup Y)}{|D|}; \quad (4)$$

$$\text{confi}R_2 = \frac{\text{sup}(X \cup Y)}{\text{sup}(X)}. \quad (5)$$

式(4)和式(5)中: X 和 Y 分别代表一个项目中不同的事务, $\text{sup}R_1$ 为规则 $R_1: X \Rightarrow Y$ 的支持度; $\text{confi}R_2$ 为规则 $R_2: X \Rightarrow Y$ 的可信度; $\text{count}(X \cup Y)$ 是 X 和 Y 并的数量; $|D|$ 是 D 所有事务的数量; $\text{sup}(X \cup Y)$ 是事务 X 并 Y 的支持度; $\text{sup}(X)$ 是事务 X 的支持度。根据公式可得,规则 R_1 的支持度即事务集中同时包含事务 X 和 Y 与所有事务集之比。规则 R_2 的可信度即为包含事务 X 的同时,出现事务 Y 的概率。

关联规则中规定,事务中的项集满足最小支持度时称为频繁项集,关联规则挖掘总体来说包括 2 个过程,即找出所有的频繁项集和由其产生的强关联规则,其中的项集的支持度满足最小支持度计数,关联规则的置信度满足最小置信度。

Apriori 算法为布尔关联规则挖掘频繁项集的原发性算法。该算法属于宽度优先算法^[22],其原理是使用逐层搜索的迭代方法,其中 $k-1$ 项集用于探索 k 项集。首先,扫描整个数据库,累计每个项的计数,找出满足最小支持度的项,得到频繁 1-项集的集合 L_1 。接着根据频繁 1-项集产生候选 2-项集,即 C_2 ,再根据 C_2 产生频繁 2-项集 L_2 ,以此循环直到无法产生新的频繁集为止。在构建候选集的同时,还需要根据先验性质“频繁项集的所有非空子集也一定是频繁的”^[23]对 C_k 进行修剪,产生对应的 L_k 。Apriori 算法的代码表达如下:

```

L1=find_frequent_1_itemset (D) //发现 1-频繁项集
for(k=2;L_{k-1}≠∅, k++)
{
C_k=apriori_gen(L_{k-1},sup_min); //根据 k-1-频繁项集产生 k-候选项集, sup_min 为最小支持度
for each t ∈ D{ //扫描记录集, 确定每个候选集的支持度
C_t=subset(C_k,t); //获得 t 所包含的候选集
For each c ∈ C_t c.count++;
}
L_k={c ∈ C_k|c.count>sup_min};
return L=∪_k L_k

```

3 结果与分析

使用 Apriori 算法对预处理后的数据进行关联规则分析。参考文献[24]，研究设定最小支持度为 10%，最小置信度为 80%，使用 Matlab 2011a 软件对立地因子和林分生长因子进行关联分析，最终得到符合条件的 175 条关联规则，由于本研究旨在挖掘影响林地立地质量和林分生长的因子间的相关规则，因此筛选提取以地位指数和林分因子作为后项且具有较高支持度、置信度的关联规则 20 条，并根据规则内容的不同，整理归结为 5 个大类(表 3)。

表 3 Apriori 算法部分关联规则结果

Table 3 Partial association rule by Apriori algorithm

规则大类	规则编号	关联规则	支持度/%	置信度/%	大类说明
类 I	1	$B_1, B_2, F_2, H_1, N_5 \rightarrow Q_1$	25.3	96.2	好立地条件为后项的规则，此类反映了较好林分立地条件下的
	2	$F_2, J_2, C_3, E_2 \rightarrow Q_1$	21.2	95.7	影响因子
	3	$C_3, J_2, H_1, D_5, D_7, G_2 \rightarrow Q_1$	22.4	87.8	
	4	$J_1, J_2, L_2, L_3, I_1 \rightarrow Q_1$	17.6	86.4	
类 II	5	$C_1, E_2, J_1, J_2 \rightarrow Q_2$	15.7	82.1	中立地条件为后项的规则，此类反映了中等林分立地条件下的
	6	$F_2, H_2, I_2, J_2, L_2 \rightarrow Q_2$	10.0	85.8	影响因子
类 III	7	$B_3, C_1, A_2, H_3 \rightarrow Q_3$	20.0	97.8	差立地条件为后项的规则，此类反映了较差林分立地条件下的
	8	$H_3, C_1, D_2, E_3 \rightarrow Q_3$	22.3	98.2	影响因子
	9	$J_4, L_1, G_1, I_3 \rightarrow Q_3$	30.1	94.5	
类 IV	10	$E_2, C_3, N_4 \rightarrow M_3$	31.4	89.6	林分生长因子为后项的规则，此类反映了影响林分生长的立地
	11	$E_3, C_1 \rightarrow M_1$	22.2	87.6	因子
	12	$I_1, E_1, Q_1 \rightarrow N_5$	28.7	92.3	
	13	$A_1, C_2, E_1, P_2 \rightarrow M_2$	17.4	86.6	
	14	$A_2, C_2, E_3, P_2 \rightarrow M_1$	21.1	83.1	
	15	$Q_1, G_2, F_1, P_3 \rightarrow O_2$	17.7	92.8	
	16	$Q_1, G_2, F_2, P_3 \rightarrow O_3$	33.1	90.2	
类 V	17	$I_3, H_2, M_1 \rightarrow P_1$	17.7	94.9	龄组为后项的规则，此类反映了不同年龄林分下的林分生长因
	18	$I_3, H_2, L_2 \rightarrow P_2$	13.8	89.6	子和立地因子的变化
	19	$I_1, J_2, L_3, O_4, O_5 \rightarrow P_3$	27.1	92.1	
	20	$J_4, J_1, L_1 \rightarrow P_5$	18.0	99.8	

各类规则的具体说明如下。规则类 I：杉木生长在海拔 400 m 以下(B_1, B_2)、黄壤(F_2)、土层厚度为厚 (H_1)的立地环境，同时平均树高在 15 m 以上 (N_5)的林分中，96.2%的林分与立地质量为好(规则 1)；立地环境为黄壤(F_2)、林下植被为草灌(J_2)、下坡位(C_3)、缓坡(E_2)的林分，有 95.7%的林分立地质量为好(规则 2)；杉木林分坡度为下(C_3)、林下植被为草灌(J_2)、土层厚度为厚(H_1)、坡向为东北(D_5)和西北(D_7)、土壤质地为壤土(G_2)的立地环境，则有 87.8%的林分立地质量为好(规则 3)；林下植被为草灌和草丛(J_1, J_2)，植被覆盖度在 31%~90%，腐殖质层为厚(I_1)的林分(L_2, L_3)中，有 86.4%的林分立地质量为好(规则 4)。规则类 II：杉木林分是上坡(C_1)和缓坡(E_2)、林下植被为草丛(J_1)或草灌(J_2)，则 82.1%的立地条件为中等(规则 5)；土壤类型为黄壤(F_2)、土层厚度为中等(H_2)、腐殖质层厚度为中等(I_2)、林下植被为草灌以及林下植被覆盖度在 31%~60%(L_2)的林分中，有 85.8%的立地质量为中等(规则 6)。规则类 III：位于海拔 400~600 m(B_3)、上坡(C_1)、丘陵地带(A_2)、土层厚度为薄(H_3)的林分中，有 97.8%的林分立地质量为差(规则 7)；土层厚度为薄(H_3)，坡位为上坡(C_1)，坡向为南坡(D_2)，坡度为斜坡(E_3)的杉木林分中，有 98.2%的立地质量为差(规则 8)。林下无植被(J_4)，植被覆盖度小于 30% (L_1)同时土壤质地为砂土(G_1)，腐殖质层为薄(I_3)的林分中，有 94.5%的立地质量为差(规则 9)。规则类 IV：生长在缓坡(E_2)，坡向为下坡(C_3)，树高在 12.1~15.0 m(N_4)的林分，有 89.6%的胸径在 15.1~20.0 cm(M_3)(规则 10)；生长在斜坡(E_3)，上坡(C_1)的林分，有 87.6%的胸径为 5.0~10.0 cm(M_1)(规则 11)；

腐殖质层为厚(I_1)、平坡(E_1)立地质量为好的林分(Q_1),有92.3%的树高位于最高等级,为15.1~18.0 m(N_5)(规则12);位于低山(A_1)中坡(C_2)平坡(E_1)的杉木中龄林,有86.6%的胸径为10.1~15.0 cm(M_2)(规则13),位于丘陵(A_2)中坡(C_2)斜坡(E_3)的杉木中龄林,有83.1%的胸径为5.0~10.0 cm(M_1)(规则14)。在好的立地质量下,黄壤(F_2)壤土(G_2)中龄林(P_2)的林分中,有92.8%的林分单位蓄积为50.0~80.0 m^3 (O_2)(规则15),而同在好的立地条件下,红壤壤土中龄林林分,有90.2%的林分单位蓄积为80.1~110 m^3 (O_3)(规则16)。规则类V:腐殖质层为薄(I_3)、土层厚度为中(H_2),胸径为5.0~10.0 cm(M_1)的林分中,有94.9%为幼龄林(规则17);腐殖质层为中(I_2)、土层厚度为中(H_2),植被覆盖度为31%~60%(L_2)的林分中,89.6%为中龄林(规则18);腐殖质层为厚(I_1)、林下植被为草灌(J_2)、蓄积在80.0~140.0 m^3 (O_4, O_5),植被覆盖度为61%~90%的林分中(L_3),有92.1%为近成熟林(规则19);林下植被为草丛(J_1)或者无植被(J_4)且植被覆盖度在30%以下(L_1)的林分中,99.8%的林分属于成熟林(规则20)。

由规则类I~III可知:影响浙西杉木用材林立地质量的立地因子有海拔、坡位、坡向、坡度级、土壤类型、土壤质地、土层厚度、腐殖质层厚度、林下植被种类和林下植被覆盖度。而地貌和林下植被高度对于立地质量好坏的影响不显著。分析发现:①立地质量随着海拔的升高而下降。究其原因是由于海拔高低决定林分生长的温度与湿度^[19],海拔升高,温度递减,湿度上升,临安地区海拔温差大,海拔较高的地区温度较低,不利于杉木林的生长。②坡位、坡向和坡度级与林分立地质量的好坏也存在一定规律,坡位越高、坡度越陡,则林分立地质量越差。这是因为坡度和坡位对林分小气候存在影响,高坡位往往处于迎风处,坡度过陡,土层越薄,迎风种植不利于杉木生长^[25]。规则还表明:东北、西北坡的杉木比南坡的立地质量好,因此说明杉木更适合生长在阴坡、半阴坡地带。③立地质量与土层厚度及腐殖质厚度成正比。有研究表明:通常土壤疏松、湿润、深厚的林分,杉木的速生期维持的时间较长^[26]。土壤中的氮磷钾等养分以及湿度会随着土层厚度的增加而增加,同时土层越厚土壤受侵蚀的程度越低^[27]。腐殖质层则能够改善土壤结构和肥力,腐殖质层越厚,土壤肥力越高,对地上植物的供养能力越充分,同时,壤土结构比砂土结构更适合杉木林的生长,这一规律也在关联规则中体现出来。④林分的立地质量与林下植被也存在着强关联。在林下植被因子中,浙西杉木用材林林下植被以草丛和草灌居多,植被覆盖度越高,立地质量越高,这是因为林下植被的增加,有利于水土保持,增加土壤渗透力及养分储存力,因此能够提高维护地力的能力。这一规律在AKPO^[28]、CACCIA等^[29]、何艺玲等^[30]的研究结果中得到证实。

规则类IV则是立地因子对林分生长的影响体现。研究表明:腐殖质层较厚的林分,树高生长良好;低山较丘陵、中坡较斜坡、下坡较上坡环境更适合杉木胸径的生长,从而间接反映出腐殖质层、地貌、坡位对林分生长环境的影响。在相同的立地条件下,与红壤相比,黄壤壤土下栽种杉木将获得更多的蓄积,这一结果与宋静^[25]在杉木生长环境的研究一致。规则IV实则是对前3类规则的另一种表达,是立地质量在林分生长中的体现。

由规则类V可知:部分林分因子随林龄的变化也呈现出一系列的变化规律,具体表现在植被覆盖度以及腐殖质层厚度2个因子中。在幼龄林阶段,林下植被覆盖度较低,林下生物量相对稀少,腐殖质层较薄,随着年龄的增长,植被覆盖度上升;在中龄林阶段以草灌为主,同时凋零物增加,腐殖质层厚度增加,林下植被覆盖度和腐殖质厚度在近成熟林阶段达到最高值;当林分逐渐成熟,林冠郁闭,灌木和草本逐渐开始消失,林下植被覆盖度下降,同时植被从高生产力转变为低生产力。这一结果与林下植被演替动态规律研究一致^[31]。

4 结论与讨论

目前,从海量数据中挖掘出潜在规则和模式是数据挖掘的基础问题。研究表明:①在一定气候区域范围内,杉木林分的立地条件主要受海拔、坡位、坡向、坡度级、土壤类型、土壤质地、土层厚度、腐殖质层厚度、林下植被种类和林下植被覆盖度的影响。根据规则及分析,在杉木造林时,应该选择低海拔、阴坡或半阴坡、坡度较缓的地带进行造林位,海拔较高处则要选择温暖地区进行种植;若要保持优质的林分立地环境,除了人为地对林地进行施肥除草等措施外,林分的林下植被也保持一定的生物多样性和植被覆盖度。此外,一般情况下,地貌也是影响立地质量的主要因子,但是本研究的数据

有限, 地貌仅体现在低山和丘陵, 因此不能比较中山、高山对杉木林生长的影响。②杉木树高的生长受土壤腐殖质层的影响较大, 而胸径的生长则受到地貌、坡位的影响较大, 因此, 若要培育中大径材的林木, 则考虑将杉木种植在低山、中下坡位的地带。在土壤选择上, 尽量选择在黄壤壤土, 土壤深厚肥沃地带进行种植, 这样将提高林分蓄积收获量。③林下植被以草灌为主, 植被覆盖度随着林分林龄的增长呈先增长后减少的变化规律。在中国人工林体系中, 往往存在林分结构简单、密度很大、林下植被不发达的现象。而关联规则表明: 林下植被多样性可以增加腐殖质厚度, 提高林分立地质量, 从而促进林木生长, 因此, 可以根据林下植被的变化规律, 适当地提高林下生境的丰富度和多样性, 也可通过适当的营林措施, 促进林下植被的发育。

以往立地质量模型中的立地因子选择往往通过主观因素而定, 应用范围小。本研究则是通过数据挖掘技术来分析评估各项立地因子之间的相互依赖关系, 从而客观地提取出与立地质量相关的因子, 构建的模型更具备科学性和实用性。

作为关联规则领域的经典算法, Apriori 算法简单易行, 能较好地完成规则提取和展示, 从而快速获取林分各类因子之间的客观规律, 具备了一定的实用性优势; 但在实际应用中, Apriori 算法需要频繁扫描数据库、产生大量候选项集从而导致时间长, 效率低, 在今后的研究中, 可对算法进行改进, 以提高数据挖掘效率。

5 参考文献

- [1] 孙岳胤, 姜文军. 沙松人工林生长与立地因子的关系[J]. 东北林业大学学报, 2004, 32(4): 86 - 87.
SUN Yueying, JIANG Wenjun. The relationships between the growth of *Abies holophylla* Maxim. plantation and the site factors [J]. *J Northeast For Univ*, 2004, 32(4): 86 - 87.
- [2] 陈淑容. 不同立地因子对楠木生长的影响[J]. 森林与环境学报, 2010, 30(2): 157 - 160.
CHEN Shurong. Effect of different site conditions on the growth of *Phoebe boumei* [J]. *J For Environ*, 2010, 30(2): 157 - 160.
- [3] 李明军, 杜明凤, 喻理飞. 贵州省杉木中龄林蓄积量与立地因子的关系研究[J]. 林业资源管理, 2014(5): 58 - 63.
LI Mingjun, DU Mingfeng, YU Lifei. Study on the relationship between the stand volume and site actors in Chinese fir forest of middle age in Guizhou Province [J]. *For Resour Manage*, 2014(5): 58 - 63.
- [4] 荀守华, 曹汉玉, 吴德军. 日本落叶松生长量与立地因子的相关分析[J]. 辽宁林业科技, 1995(2): 35 - 37.
XUN Shouhua, CAO Hanyu, WU Dejun. Correlation analysis between the increment and site factors of Japan larch [J]. *J Shandong Norm Univ Natu Sci*, 1995(2): 35 - 37.
- [5] 吴云飞, 郎南军, 李甜江, 等. 云油茶的生长指标及其立地因子的相关性分析与主导性分析[J]. 西部林业科学, 2014, 43(2): 79 - 84.
WU Yunfei, LANG Nanjun, LI Tianjiang, et al. Growth indexes of a new variety of *Camellia oleifera* in Yunnan and relativity and dominant analysis of its site factors [J]. *J West China For Sci*, 2014, 43(2): 79 - 84.
- [6] BERKHIN P. A survey of clustering data mining techniques [J]. *Grouping Multidimensional Data*, 2006, 43(1): 25 - 71.
- [7] HAN Jiawei, KAMBER M, PEI Jian. *Data Mining: Concepts and Techniques* [M]. 3 ed. Beijing: China Machine Press, 2012: 1 - 18.
- [8] ASHRAF M I, ZHAO Z, BOURQUE P A, et al. Integrating biophysical controls in forest growth and yield predictions with artificial intelligence technology [J]. *Can J For Res*, 2013, 43(12): 1162 - 1171.
- [9] YAN Fei, GONG Yinxi, FENG Zhongke. Combination of artificial neural network with multispectral remote sensing data as applied in site quality evaluation in Inner Mongolia [J]. *Croatian J For Eng*, 2015, 36(2): 307 - 319.
- [10] PAPPU J S M, VIJAYAKUMAR G K, RAMAMURTHY V. Artificial neural network model for predicting production of *Spirulina platensis*, in outdoor culture [J]. *Bioresour Technol*, 2013, 130(2): 224 - 230.
- [11] CHENG Tao, WANG Jiaqiu. Integrated spatio-temporal data mining for forest fire prediction [J]. *Trans Gis*, 2008, 12(5): 591 - 611.
- [12] POURTAGHI Z S, POURGHASEMI H R, ARETANO R, et al. Investigation of general indicators influencing on for-

- est fire and its susceptibility modeling using different data mining techniques [J]. *Ecol Indic*, 2016, **64**: 72 – 84.
- [13] 王阆, 彭世揆, 余光辉. Apriori 算法在森林资源二类调查数据分析中的应用[J]. 南京林业大学学报(自然科学版), 2006, **30**(3): 63 – 66.
WANG Tian, PENG Shikui, SHE Guanghui. Application of Apriori algorithm on data analyses of forest inventory [J]. *J Nanjing For Univ Nat Sci Ed*, 2006, **30**(3): 63 – 66.
- [14] LIN Jiexiong, HUANG Zhan. An improved Apriori algorithm based on array vectors [J]. *Comput Appl Soft-ware*, 2005, **28**(5): 268 – 271.
- [15] NAJADAT H M, AL-MAOLEGI M, ARKOK B. An improved Apriori algorithm for association rules [J]. *Int Res J Comput Sci Appl*, 2013, **1**(1): 1 – 8.
- [16] 周国模, 郭仁鉴, 韦新良, 等. 浙江省杉木人工林生长模型及主伐年龄的确定[J]. 浙江林学院学报, 2001, **18**(3): 219 – 222.
ZHOU Guomo, GUO Renjian, WEI Xinliang, *et al.* Growth model and cutting age of Chinese fir planted forest in Zhejiang Province [J]. *J Zhejiang For Coll*, 2001, **18**(3): 219 – 222.
- [17] 李永宁, 张宾兰, 秦淑英, 等. 郁闭度及其测定方法研究与应用[J]. 世界林业研究, 2008, **21**(1): 40 – 46.
LI Yongning, ZHANG Binlan, QIN Shuying, *et al.* Review of research and application of forest canopy closure and its measuring methods [J]. *World For Res*, 2008, **21**(1): 40 – 46.
- [18] 郭艳荣, 吴保国, 刘洋, 等. 立地质量评价研究进展[J]. 世界林业研究, 2012, **25**(5): 47 – 52.
GUO Yanrong, WU Baoguo, LIU Yang, *et al.* Research progress of site quality evaluation [J]. *World For Res*, 2012, **25**(5): 47 – 52.
- [19] 杜纪山, 唐守正, 王洪良. 天然林分生长模型在小班数据更新中的应用[J]. 林业科学, 2000, **36**(3): 52 – 58.
DU Jishan, TANG Shouzheng, WANG Hongliang. Application of natural stand growth models in subcompartment data update [J]. *Sci Silv Sin*, 2000, **36**(3): 52 – 58.
- [20] CARMEAN W H, LENTHALL D J. Height-growth and site-index curves for jack pine in north central Ont. [J]. *Can J For Res*, 1989, **19**(2): 215 – 224.
- [21] GUO Yanrong, HAN Yanyun, WU Baoguo, *et al.* Study on modelling of site quality evaluation and its dynamic update technology for plantation forests [J]. *Nat Environ Pollut Technol*, 2013, **12**(4): 591 – 597.
- [22] ORLANDO S, PALMERINI P, PEREGO R. *Enhancing the Apriori Algorithm for Frequent Set Counting* [M]. Berlin: Springer Berlin Heidelberg, 2001: 71 – 82.
- [23] 韩家炜, MICHELINE K, 裴健. 数据挖掘概念与技术[M]. 北京: 机械工业出版社, 2012: 2 – 17.
- [24] 王霓虹, 高萌, 李丹, 等. 基于 C5.0 与 Apriori 算法的森林生物量等级评价与因子关联分析[J]. 中南林业科技大学学报, 2015, **35**(3): 1 – 6.
WANG Nihong, GAO Meng, LI Dan, *et al.* Evaluation on forest biomass grade and analyses on factor correlations by using C5.0 and Apriori algorithm [J]. *J Cent South Univ For Technol*, 2015, **35**(3): 1 – 6.
- [25] 宋静. 浅议杉木生长与环境条件的关系[J]. 中国西部科技, 2008, **7**(11): 54 – 55.
SONG Jing. Discussion of the relationship between Chinese fir growth and environmental conditions [J]. *Western China Technol*, 2008, **7**(11): 54 – 55.
- [26] 迟健. 杉木速生丰产优质造林技术[M]. 北京: 金盾出版社, 1996.
- [27] MONSERUD R A, MARSHALL J D. Time-series analysis of ^{13}C from tree rings. I. Time trends and autocorrelation [J]. *Tree Physiol*, 2001, **21**(15): 1087 – 1102.
- [28] AKPO L E. Phenological interactions between tree and understory herbaceous vegetation of a Sahelian semi-arid Savanna [J]. *Plant Ecol*, 1997, **131**(2): 241 – 248.
- [29] CACCIA F D, BALLARÉ C L. Effects of tree cover, understory vegetation, and litter on regeneration of Douglas-fir (*Pseudotsuga menziesii*) in southwestern Argentina [J]. *Can J For Res*, 1998, **28**(5): 683 – 692.
- [30] 何艺玲, 傅懋毅. 人工林林下植被的研究现状[J]. 林业科学研究, 2002, **15**(6): 727 – 733.
HE Yiling, FU Maoyi. Review of studies on understorey of plantations [J]. *For Res*, 2002, **15**(6): 727 – 733.
- [31] ALABACK P B. *Plant Succession Following Logging in the Sitka Spruce-western Hemlock Forests of Southeast Alaska: Implications for Management* [R]. Washington: USDA Forest Service General Technical Report, 1984.