

# 基于约束满足和贝叶斯网络的大规模定制产品配置方法

苏依拉, 李慧旻, 王 斐

(内蒙古工业大学 信息工程学院, 呼和浩特 010051)

**摘 要:** 针对在产品配置过程中忽略了用户兴趣这一不足, 提出了一种基于约束满足问题 (constraint satisfaction problem, CSP) 与贝叶斯网络 (Bayesian network, BN) 二者结合的产品配置方法. 在 CSP 与 BN 的理论基础上, 建立了基于 CSP 与 BN 的配置模型, 给出了求解的具体方法. 最后以组装电脑为例, 开发了组装电脑产品的配置系统, 该系统验证了这种配置方法能对配置结果进行优化, 具有一定的可行性和有效性.

**关键词:** 大规模定制; 产品配置; 约束满足问题 (CSP); 贝叶斯网络 (BN)

中图分类号: TP 399

文献标志码: A

文章编号: 0254-0037(2015)07-1005-07

doi: 10.11936/bjutxb2014100074

## Mass Customized Product Configuration Method Based on Constraint Satisfaction Problem and Bayesian Networks

SU Yi-la, LI Hui-min, WANG Fei

(College of Information Engineering, Inner Mongolia University of Technology, Hohhot 010051, China)

**Abstract:** To address the deficiency of ignoring the interests of users in the process of product configuration, a combined product configuration method based on the CSP and Bayesian Network (BN) was proposed in this paper. On the basis of CSP and BN theory, a product configuration model was established based on CSP and BN, and the specific method of solution was given, to assemble the computer as an example, a configuration system of assembly computer product was developed so as to verify the optimized configuration result, and it shows good feasibility and effectiveness.

**Key words:** mass customization; product configuration; constraint satisfaction problem; Bayesian network

随着科学技术的发展, 制造业面临着越来越激烈的竞争和挑战, 产品生产周期的大幅缩短、产品需求量的急剧增加、用户需求的个性化和多样化等等. 在这样的趋势下, 制造业的生产模式由传统的大规模生产逐渐过渡到大规模定制生产. 大规模定制<sup>[1]</sup>的提出使大规模生产与用户的个性化需求二者有机的结合在一起.

产品配置技术是实现大规模定制的关键技术手段, 它是人工智能领域内的一个成功应用. 产品配置又包括了诸多方法, 其中基于约束满足问题 (constraint satisfaction problem, CSP) 的产品配置方法虽然具有较高的求解效率<sup>[2]</sup>, 但它具有较弱的相关性, 即在产品配置过程中没有充分地考虑用户兴趣的因素, 产生了许多不符合用户需求的结果. 对

收稿日期: 2014-10-28

基金项目: 国家自然科学基金资助项目(61363052); 内蒙古自然科学基金资助项目(2012MS0904)

作者简介: 苏依拉(1964—), 男, 教授, 主要从事人工智能、机器学习、数据挖掘、计算机网络方面的研究, E-mail: suyila@tsinghua.org.cn

用户来说,配置的求解过程他们并不关心,而在乎的是能否在有限的时间内真正获得满足自己需求的配置结果.所以在产品配置的过程中如何获得用户的偏好,如何得到用户真正满意的配置结果是当前产品配置领域研究的重点.

针对上述问题,本文提出了一种将贝叶斯网络(Bayesian network, BN)与约束满足问题(constraint satisfaction problem, CSP)二者相结合的产品配置方法,包括配置知识表示和配置求解这2个过程.其中BN相当于一个中间件,由用户提供的约束推理出新的约束,这种新的约束集合就是用户潜在的兴趣信息,同时也是对约束满足问题的一种扩展.

## 1 相关概念

### 1.1 CSP

一个约束通常是指一个包含若干变量的关系表达式,用以表示这些变量所必须满足的条件.人工智能和其他计算机科学领域中的大量问题都可归结为CSP,例如定性推理、基于模型的诊断、自然语言理解、景物分析、任务调度、系统配置、科学实验规划、机械与电子设备的设计与分析等.

形式化地说,一个CSP是由 $X = \{X_1, X_2, \dots, X_n\}$ 、 $C = \{C_1, C_2, \dots, C_m\}$  2个集合定义的. $n$ 个变量有一个非空的有限值域集合 $D = \{D_1, D_2, \dots, D_j\}$ ,所以一个CSP问题也可表示为一个三元组 $W = (X, D, C)$ ,其中 $X$ 表示变量的集合, $D$ 表示一个非空的有限值域集合, $C$ 表示约束集合,每个约束 $C_i$ 包括一些变量的子集,并指定了这些子集的值之间允许进行的合并.问题的一个状态是由对一些或全部变量的一个赋值定义的,一个不违反任何约束条件的赋值称作合法或相容的赋值.每个变量都参与的赋值称为一个完全赋值,而CSP的解就是满足所有约束条件的完全赋值<sup>[3]</sup>.

### 1.2 BN

BN又称为信念网络(belief networks)、概率网络(probabilistic networks)或因果网络(causal networks).它是目前人工智能领域内表达不确定性知识并能进行推理的最有效的理论模型之一,也是数据挖掘技术中常用的方法之一<sup>[4]</sup>.BN是一种用来表示变量间连接概率的有向图形模式,它提供了一种自然的表示因果信息的方法,用来发现数据间的潜在关系.BN用节点表示变量,用有向边表示变量间的依赖关系.其详细的描述如下:

1) 一个随机变量集组成网络节点.变量可是

离散的或者连续的.

2) 一个连接节点对的有向边或箭头组成边的集合.如果存在从节点 $X$ 指向节点 $Y$ 的有向边,则称 $X$ 是 $Y$ 的一个父节点.

3) 每个节点 $X_i$ 都有一个条件概率分布 $P(X_i | \text{Parents}(X_i))$ ,表示其父节点对该节点的影响.

4) 图中不存在有向环,因此是一个有向无环图(directed acyclic graph, DAG)<sup>[5]</sup>.

目前对BN的研究主要是推理和学习,其中BN的推理是应用BN对问题进行求解的过程<sup>[6]</sup>,包括精确推理和近似推理,学习包括网络拓扑结构的学习和参数学习.BN不仅是一种基于贝叶斯概率理论的图形化建模工具,而且它具有不确定性推理能力等特点,因此,可用其解决产品配置领域中配置推理优化问题.

## 2 基于CSP与BN的配置知识表示

### 2.1 配置模型的建立

传统的CSP配置模型是一个四元组 $(X, D, C, C_u)$ , $C_u$ 为用户的约束条件集合;一般情况下,在CSP中,给出的约束条件越多,得到的结果就越精确.同样,在基于CSP的产品配置系统中,为了让配置结果更加满足于用户的需求,应对 $C_u$ 进行扩展,也是对CSP的一种扩展,将用户潜在的偏好集合作为扩展约束更能体现配置的准确性.因此,本文提出的基于CSP与BN的配置模型是一个五元组 $(X, D, C, C_u, C_e)$ ,其中 $C_e$ 为用户的偏好集合.

一般情况下,在对CSP进行求解时,对给定赋值后的变量个数越多,问题的解越具体.该模型相当于扩展了CSP的变量空间, $C_e$ 作为扩展约束,对这个五元组的求解也是对CSP的求解.

### 2.2 用户的BN模型的构建

用户的偏好通过BN模型进行推理得到,建模的过程包括了贝叶斯网的结构学习和参数学习,其中参数学习也是对网络节点条件概率的学习.本文根据产品配置领域的特点,直接采用了产品的逻辑结构作为BN拓扑结构,以组装电脑为例,如图1所示.该结构为树形,是一个特殊的BN结构,将产品逻辑结构中的节点作为BN的节点,这些节点也作为用户的偏好集合,有向边则直观地表示了节点之间的关系.

在这个结构中包含了3种关系<sup>[7]</sup>:

1) Is\_A关系,表示逻辑部件在不同抽象层次上的泛化关系,如Opteron64是一种(Is\_A)CPU;

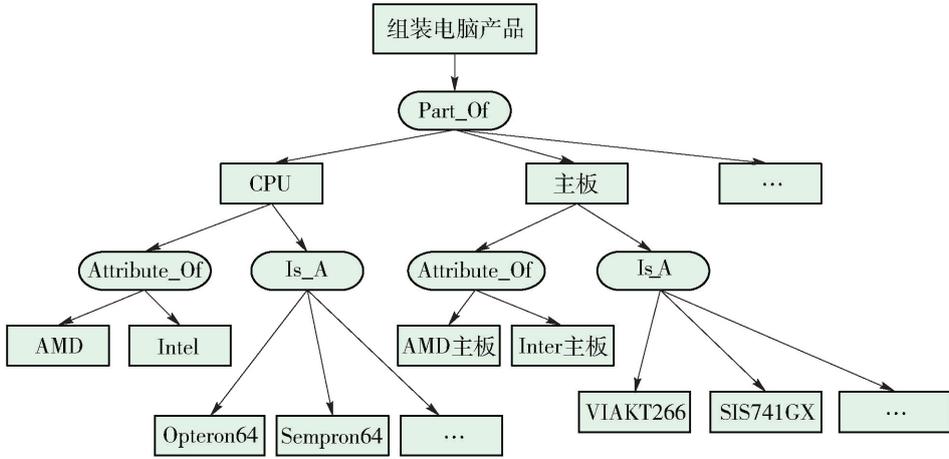


图 1 组装电脑的逻辑结构

Fig. 1 Logical structure of assembled computer

2) Part\_Of 关系, 表示逻辑部件间的聚合关系, 即逻辑部件与它的直接子部件间形成的聚合关系;

3) Attribute\_Of 关系, 表示逻辑部件的属性关系, 如 CPU 的属性型号有 AMD 和 Intel 两种.

采用这种树形 BN 结构的优势是在进行推理时, 能大大提高推理的速度. 因为在非树形结构的 BN 上进行推理一般都是 NP 问题, 但对于树形结构的 BN, 许多推理算法都可在线性时间内完成. 这不仅简化了推理过程, 也提高了产品的配置效率, 实现了对用户的快速响应.

在建好 BN 结构后就要对网络中各个节点的条件概率表进行学习. 条件概率表的学习是通过统计数据库中的数据来实现的.

本文学习的数据样本取自一段时间内单独基于 CSP 的产品配置系统中的 1 000 条用户配置结果, 它是一个完整的实例数据库, 所以数据是完备的, 通过样本统计法和贝叶斯方法就可计算出来. 学习的过程分为 2 个步骤.

1) 配置结果格式化

将每条配置结果记录成如表 1 所示的格式, 其中记录中的每个变量都是布尔值 {0, 1}, 0 表示用户未购组件, 1 表示用户已购组件.

2) 统计数据库学习条件概率

得到格式化的配置结果后, 还需得出统计数据库学习条件概率. 根据贝叶斯公式, 例如: Sempron64 的条件概率  $P(\text{Sempron64} = 1 | \text{AMD} = 1)$  的计算方法为

$$P(\text{Sempron64} = 1 | \text{AMD} = 1) = \frac{P(\text{Sempron64} = 1 \wedge \text{AMD} = 1)}{P(\text{AMD} = 1)}$$

式中:  $P(\text{Sempron64} = 1 \wedge \text{AMD} = 1)$  为 Sempron64 与 AMD 同时取 1 的记录总数;  $P(\text{AMD} = 1)$  为 AMD 取 1 的记录总数. 所以通过对数据库进行统计就能求出  $P(\text{Sempron64} = 1 | \text{AMD} = 1)$  的条件概率. 同理, 用该方法能求出 Sempron64 节点条件概率表的其他项以及其余节点的条件概率表.

表 1 用户历史配置信息的记录格式

Table 1 Record format of history configuration for user configuration

序号	AMD	Sempron64	HY	Seagate	Kingston
1	1	0	1	1	0
2	1	1	0	1	1
3	0	1	1	0	1

通过上述方法就确定了用户的 BN 的拓扑结构和参数, 即确定了用户的 BN 模型, 那么下一步就是配置求解的过程, 即求解配置模型五元组  $(X, D, C, C_u, C_e)$  中的  $C_e$  与配置推理.

3 基于 CSP 与 BN 的配置求解

3.1 配置求解流程

本文以组装电脑产品为研究对象, 研究了适合组装电脑的产品配置知识表示方法, 根据其结构和配置模型的特点, 采用了基于 CSP 和 BN 相结合的产品配置求解方法. 具体的求解步骤如下:

1) 用户给出需求约束集合  $C_u$ .

2) 通过建立好的用户 BN 模型根据  $C_u$  进行计算推理, 推理过程中将  $C_u$  作为证据节点, 由此推理除证据节点外其他节点的后验概率.

3) 设定推荐阈值  $\varepsilon$ , 将后验概率大于  $\varepsilon$  的节点

与用户需求约束进行一致性检验,将满足条件的节点集作为扩展约束  $C_e$ ,得到的  $C_e$  就是用户的偏好集合. 如果没有任何节点满足推荐阈值或者满足推荐阈值的节点都没有通过一致性检验,那么直接进行 CSP 四元组  $(X, D, C, C_u)$  的推理.

4) 将  $C_e$  与  $C_u$  共同代入基于 CSP 与 BN 二者结合的配置模型  $(X, D, C, C_u, C_e)$  进行配置推理.

5) 对该五元组的求解也是对 CSP 的求解,最后将求解的结果信息反馈给用户.

从配置流程中可看出,配置的求解主要包括 2 个过程:用户偏好的获得,即对 BN 模型进行概率推理;对五元组  $(X, D, C, C_u, C_e)$  的求解,即对 CSP 的求解.

配置求解的流程如图 2 所示.

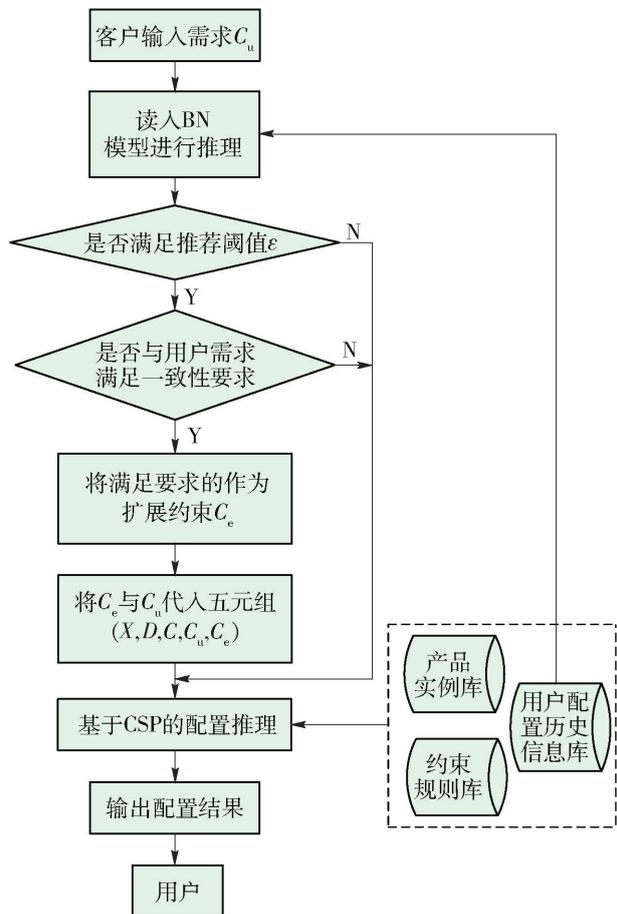


图 2 基于 CSP 与 BN 的配置流程

Fig. 2 Flow chart of configuration based on CSP and BN

### 3.2 BN 的概率推理过程

BN 推理的目的是求出用户潜在的兴趣节点集合,即用户的偏好,所以由用户的配置历史信息建立了用户的 BN 模型,2.2 节中已经给出了建模的方法,但是手动创建 BN 模型是一个很复杂的过程,针

对这样的问题,许多关于 BN 建模的软件被开发出来,它们的侧重点都是不同的,其中,基于 Matlab 的 BN 工具箱 BNT (Bayesian networks toolbox) 是目前功能最全的软件. 它是由 Kevin 基于 Matlab 语言开发的开源软件包,主要是关于 BN 学习,而且也包含了很多有关 BN 学习的底层基础函数库,其功能十分丰富,不仅支持精确推理和近似推理,而且也支持拓扑结构的学习和参数学习以及静态模型和动态模型<sup>[8]</sup>. 本文采用 BN 工具包 BNT 去建模,能实现对用户需求约束的扩展. 在 BNT 中,采用矩阵的形式来表示 BN,若节点  $i$  到节点  $j$  由一条有向边相连,则对应矩阵中  $(i, j)$  值为 1, 否则为 0. 下面为 BN 的矩阵形式. BNT 的推理过程如图 3 所示.

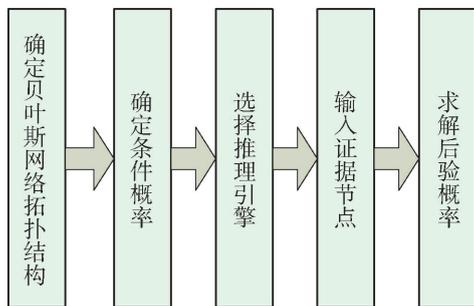


图 3 BNT 的推理过程

Fig. 3 Reasoning process of BNT

本文应用 BNT 建立用户的 BN 模型,首先创建 BN 的结构. 网络节点的条件概率是通过数据样本进行学习,该样本取自一段时间内基于 CSP 方法配置系统中的用户历史配置记录. 记录中每个字段都是产品逻辑结构中的各个节点,对于这些节点用户体现的购买行为有 2 种:1 表示未配置此部件,2 表示配置过此部件,所以将每条记录都用 1 和 2 的形式来表示. 通过 BNT 进行结构学习后如图 4 所示,包含了 37 个节点,每个节点变量的取值是  $\{1, 2\}$ ,有向边体现了节点之间的关系.

在 BNT 中,本文将数据样本放入一个 traindata.txt 文本文件里作为训练样本,每一行为一个用户的配置实例用 1 和 2 来表示,包括了所有节点的记录格式,节点之间用逗号隔开,学习的方法采用的是贝叶斯估计法,在 BNT 中的学习语句是 bayes\_update\_params().

确定了 BN 的结构和参数即确定了 BN 模型,接下来就可通过该模型进行概率推理. 本文所采用的概率推理算法属于 BN 精确推理算法,结合产品配置系统在线推荐的特点,需要将这些推荐信息及时

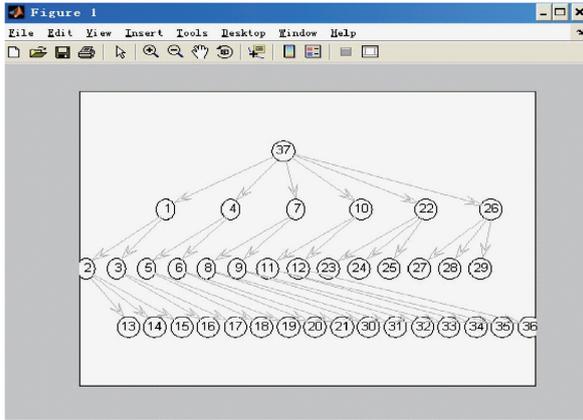


图 4 组装电脑 BN 结构在 BNT 中的表示

Fig. 4 Bayesian network structure of assembled computer in BNT

反馈给用户,因此,对于 BN 的推理速度要求较高. 所以本文采用的是联合树算法,因为该算法不仅具有推理结果精确、高效等特点,在推理速度方面是目前最快的,也是 BN 精确推理算法中应用最为广泛的.

BNT 工具箱就采用了引擎机制,不同的引擎根据不同的算法来完成模型转换、细化和求解,其中就包括了联合树推理引擎:

```
jtree_inf_engine();
```

那么推理的过程如图 5 所示.

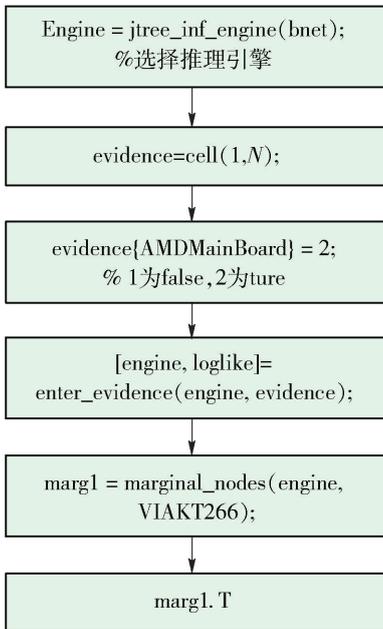


图 5 推理过程

Fig. 5 Reasoning process

这里输入的证据节点为 AMDMainBoard,推理 VIAKT266 的后验概率及其边缘分布分别如图 6、7

所示.

```
The probability of VIAKT266:
ans =

    0.6004
    0.3996

>>
```

图 6 VIAKT266 后验概率

Fig. 6 Posterior probability of VIAKT266

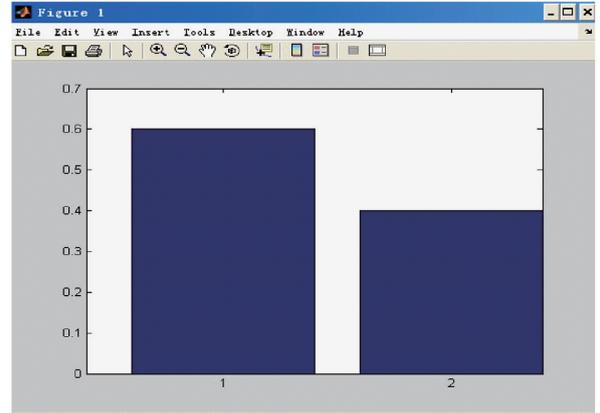


图 7 VIAKT266 边缘分布

Fig. 7 Marginal distribution of VIAKT266

### 3.3 五元组 $(X, D, C, C_u, C_e)$ 的求解

这个过程也是对 CSP 求解的过程,本文采用的是一种改进的回溯法前向检验算法. 回溯法是搜索算法最常用的算法<sup>[9]</sup>,对回溯法的改进可通过约束传播来实现,本文所采用的方法是前向检验(forward checking, FC),它是前看策略中一种比较简单而有效的方法<sup>[10]</sup>. 这种算法与传统的回溯搜索相对一致性的要求更高,在求解 CSP 时不但可减少对新变量赋值时的搜索空间,而且能判断当前的部分赋值能否被扩展为一个解,提前发现无解的状态,避免无效的搜索. 在前向检验 FC 和联合树算法的基础上,基于 CSP 与 BN 配置求解的具体步骤如下:

1) 用户输入约束条件对配置变量进行赋值,将这些赋值变量传输到 BN 模型进行概率推理,判断除用户赋值的变量外其余变量后验概率是否满足推荐阈值. 如果都不满足,则将扩展约束变量置为空;取出满足条件的变量进行一致性检验,取出满足条件的变量作为扩展约束,如果都不满足,将扩展约束变量置为空.

2) 将用户输入变量和推理得到的扩展约束变量作为过去变量  $X_1$  (已赋值的配置变量集合),如果未来变量  $X_t$  (未赋值的配置变量集合) 为空集,输出结果,程序结束;否则从未来变量的集合中选择一个

未来变量作为当前变量  $X_c$  (正在赋值的配置变量集合).

3) 从当前变量  $X_c$  的值域  $D_c$  中选择一个值  $d_c$  赋给  $X_c$ , 并将  $D_c$  从  $X_c$  的域值中删掉, 然后更新当前变量  $X_c$  的值域.

4) 对被赋值的当前变量  $X_c$  与其相关的约束集合进行一致性检验, 如果某一个未来变量的值域变为空集, 那么当前变量  $X_c$  则与这个未来变量不相容, 程序发生回溯, 即将问题状态恢复到约束检验前的状态, 转到步骤 5); 如果未发现不相容变量, 则更新当前所有变量的值域, 并转到步骤 2).

5) 如果当前变量  $X_c$  的值域为空, 那么程序回溯到上次赋值的过去变量, 并将当前变量和未来变量的值域恢复到该过去变量上次赋值前的状态, 同时将该过去变量作为当前变量, 转到步骤 3).

### 4 基于 CSP 与 BN 的产品配置实例

本文结合 CSP 与 BN 的方法开发了组装电脑产品的配置系统, 将这些理论与方法在系统中得以实现, 并验证了它的有效性. 其系统框架如图 8 所示.

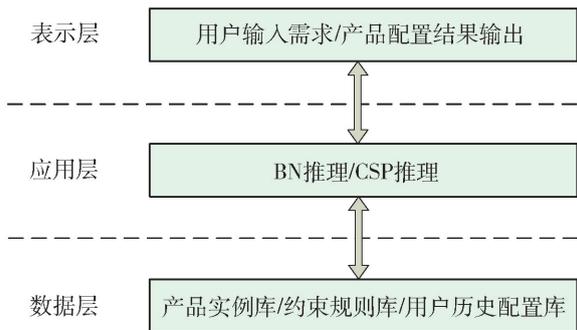


图 8 系统的框架结构

Fig. 8 System framework

在表示层, 用户的需求可通过下拉列表进行选择, 也是对各个变量进行赋值的过程. 为满足不同类型的用户, 在选择上包括了特性需求选择和具体零部件的选择 2 种. 例如用户选择了 CPU 种类、主板种类、硬盘型号、显示器尺寸这 4 个选项, 那么根据这 4 个约束进行 CSP 推理得到的配置结果是 36 条. 同样选择这 4 条约束, 应用基于 CSP 与 BN 的方法, 在配置求解时, 对于结合 CSP 与 BN 二者的配置推理机制, 首先设置推荐阈值, 在本系统中设置该值为 0.5, 根据用户输入的 4 个约束条件, 将它们作为用于贝叶斯网推理的证据节点. 经 Matlab 工具箱 BNT 推理取出满足阈值的节点, 同时判断这些节点与证据节点是否满足一致性, 即是否满足与证据节

点相连的约束规则, 将不一致的节点删除, 那么将概率大于 0.5 的剩余节点作为扩展约束进行 CSP 推理. 点击 CSP 与 BN 推理按钮进行配置求解, 配置结果如图 9 所示, 共 8 条记录. 其中满足扩展约束条件的节点概率在 Matlab 中的显示如图 10 所示, 经过用户 BN 模型的推理, 得到的符合条件的节点有 Sempron64、Opteron64、GeForce、SIS741GX 和 Nvidia. 对于每个节点的 2 个概率值, 在进行约束推荐时是判断第 2 个值是否满足推荐阈值, 如节点 Sempron64 作为扩展约束的概率为 0.8968.

```

Command Window
The probability of Sempron64:
ans =
    0.1032
    0.8968

The probability of Opteron64:
ans =
    0.3724
    0.6276

The probability of GeForce:
ans =
    0.3864
    0.6136

The probability of SIS741GX:
ans =
    0.4109
    0.5891

The probability of Nvidia:
ans =
    0.4451
    0.5549
>>

```

图 9 满足推荐阈值的概率结果

Fig. 9 Probability results of meeting the recommended threshold

配置结果编号	CPU	CPU种类	主板	主板种类	显示器	显示器尺寸	硬盘	硬盘接口	内存	显卡
1	Sempron64	AMD系列	SIS741GX	AMD主板	BEHQ	17寸显示器	Seagate	SATA	KingMax	GeForce
2	Sempron64	AMD系列	SIS741GX	AMD主板	TCL	17寸显示器	Seagate	SATA	KingMax	GeForce
3	Sempron64	AMD系列	SIS741GX	AMD主板	BEHQ	17寸显示器	Seagate	SATA	KingMax	Nvidia
4	Sempron64	AMD系列	SIS741GX	AMD主板	TCL	17寸显示器	Seagate	SATA	KingMax	Nvidia
5	Opteron64	AMD系列	SIS741GX	AMD主板	BEHQ	17寸显示器	Seagate	SATA	KingMax	GeForce
6	Opteron64	AMD系列	SIS741GX	AMD主板	TCL	17寸显示器	Seagate	SATA	KingMax	GeForce
7	Opteron64	AMD系列	SIS741GX	AMD主板	BEHQ	17寸显示器	Seagate	SATA	KingMax	Nvidia
8	Opteron64	AMD系列	SIS741GX	AMD主板	TCL	17寸显示器	Seagate	SATA	KingMax	Nvidia

图 10 基于 CSP 和 BN 求解的配置结果

Fig. 10 Results of configuration based on CSP and BN

## 5 结论

1) 在同样的用户需求约束条件下,使用基于 CSP 与 BN 的配置方法得到的配置结果明显优于传统基于 CSP 的配置方法,在数量上就已经精简许多。

2) 使用基于 CSP 与 BN 的配置方法在配置求解时不仅能减少不相关项,而且提高了配置的准确性。

3) 这种配置方法能对配置结果进行优化,具有一定的可行性和有效性。

### 参考文献:

- [1] STANLY D. From future perfect: mass customizing [J]. Planning Review, 1989, 17(2): 16-21.
- [2] GELLE E, WEIGEL R. Interactive configuration using constraint satisfaction techniques [C] // Proc Artificial Intelligence and Manufacturing Research Planning Workshop. Washington DC: AAAI Press, 1996: 37-44.
- [3] 李伟, 刘光复, 张孟青. 一种基于约束满足问题的产品配置方法[J]. 成组技术与生产现代化, 2004, 21(2): 50-54.  
LI Wei, LIU Guang-fu, ZHANG Meng-qing. A product configuration method based on constraint satisfaction problems [J]. Group Technology & Production Modernization, 2004, 21(2): 50-54. (in Chinese)
- [4] JENSEN F. Bayesian networks and decision graphs [M]. New York: Springer-Verlag, 2001.
- [5] 徐从富, 李石坚, 王金龙. 机器学习研究与应用新进展 [M]. 杭州: 浙江大学人工智能研究所, 2006: 8.
- [6] 刘俊娜. 贝叶斯网络推理算法研究[D]. 合肥: 合肥工业大学计算机与信息学院, 2007.  
LIU Jun-na. A Bayesian network inference algorithm[D]. Hefei: School of Computer and Information, HeFei University of Technology, 2007. (in Chinese)
- [7] 吴先超, 吕晓枫, 孙吉贵. 基于广义产品结构的配置和集成研究[J]. 计算机集成制造系统, 2006, 12(6): 868-875.  
WU Xian-chao, LÜ Xiao-feng, SUN Ji-gui. The research on configuration and integration based on generalized product structure [J]. Computer Integrated Manufacturing Systems, 2006, 12(6): 868-875. (in Chinese)
- [8] 吴良清. 基于 Bayes 网的软件可靠性模型研究与系统设计[J]. 电子工程师, 2007, 33(5): 41-48.  
WU Liang-qing. Research on software reliability and system design base on Bayes network [J]. Electronics Engineer, 2007, 33(5): 41-48. (in Chinese)
- [9] DECHER Rina. Constraint processing [M]. Burlington: Morgan Kaufmann, 2003: 9-10.
- [10] KASK K, DECHTER R, GOGATE V. Counting-based look-ahead schemes for constraint satisfaction [M]. Burlington: Morgan Kaufinann, 2003: 6-9.

(责任编辑 吕小红)