

# 混合约束长度满足问题的分簇相变

金 迪

上海理工大学理学院, 上海

收稿日期: 2025年3月3日; 录用日期: 2025年4月4日; 发布日期: 2025年4月18日

## 摘 要

随机约束满足问题解空间的结构与问题的难解性之间具有紧密的内在联系。针对随机约束满足问题中固定约束长度模型与实际场景中非均匀约束的差异性, 提出了具有混合约束长度的 $RB^{mix}$ 模型。基于统计物理的序参量理论, 利用解对间Hamming距离演化分析与矩方法相结合, 揭示了混合约束满足问题模型解空间的分簇相变。结果表明, 当控制参数达到临界阈值时, 解空间发生从复本对称态到统计可分簇态的相变, 其特征表现为解集内部涌现指数级规模的子簇, 且相邻子簇间存在超Hamming距离的能量壁垒。这揭示了分簇相变过程中解空间连通性的突变规律为基于相变特征的组合优化算法设计提供了理论依据。

## 关键词

约束满足问题,  $RB^{mix}$ 模型, 混合约束, 分簇相变

# The Clustering Phase Transition of Random Mixed Constraint Satisfaction Problem

Di Jin

College of Science, University of Shanghai for Science and Technology, Shanghai

Received: Mar. 3<sup>rd</sup>, 2025; accepted: Apr. 4<sup>th</sup>, 2025; published: Apr. 18<sup>th</sup>, 2025

## Abstract

There exists a profound intrinsic connection between the structure of the solution space and the computational hardness of random constraint satisfaction problems (CSPs). To address the discrepancy between fixed constraint length models in random CSPs and the non-uniform constraints in real-world scenarios, the  $RB^{mix}$  model with mixed constraint lengths has been proposed. Leveraging the order parameter theory from statistical physics and combining the analysis of Hamming distance evolution between solution pairs with the moment method, the clustering phase transition in the solution space of the mixed constraint satisfaction problem model has been revealed. The

results demonstrate that when the control parameter reaches a critical threshold, the solution space undergoes a phase transition from a replica-symmetric state to a statistically clustered state. This transition is characterized by the emergence of exponentially many sub-clusters within the solution set, accompanied by energy barriers of ultra-Hamming distances between adjacent sub-clusters. This reveals the abrupt change in the connectivity of the solution space during the clustering phase transition, providing a theoretical foundation for the design of combinatorial optimization algorithms based on phase transition characteristics.

## Keywords

Constraint Satisfaction Problem,  $RB^{\text{mix}}$  Model, Mixed Constraint, Clustering Phase Transition

Copyright © 2025 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

## 1. 引言

约束满足问题(Constraint Satisfaction Problems, CSP)是计算机科学、数学、物理学和工程学等多个领域中的一个重要研究主题,广泛应用于资源分配、逻辑推理、蛋白质结构预测和化学分子设计等实际问题中。CSP 问题的核心在于找到一组变量赋值,同时能够满足所有约束条件。由于其普遍性和复杂性,CSP 问题在计算复杂性理论中占据重要地位,尤其是 NP-完全问题的研究中[1],通过对约束满足问题的深入研究,理论上可以帮助揭示计算复杂性的本质,还可以为设计高效算法和优化资源分配提供重要工具。近年来,随着对 CSP 的深入研究,解空间结构逐渐成为理解问题复杂性、算法性能和相变现象的关键。当控制参数变化时,解空间也会随之发生不同的结构性变化,表现出数值实验中就是当控制参数小于某一数值时,问题能找到可满足解;而大于这一数值时,问题实例的可满足概率会出现从 1 到 0 的突变,且在这个临界值附近求解难度到达最大值。这一有趣现象引发了人们对平均计算复杂与解空间结构之间存在关系的猜测。

最早的 CSP 研究基于四个基础模型为 A、B、C、D 模型[2]-[4],但是 Achlioptas 等人指出,当问题规模增大时,这些基础模型都表现出渐近平凡不可满足性[5]。为克服这一缺陷,研究学者通过改变控制参数变化的方式,提出了 RB 模型,即约束数目和取值域规模都随变量个数呈指数增长。然而, RB 模型的随机实例由固定长度约束组成,对于生活中很多实际问题来说,约束的形式是变化的。基于 RB 模型,许多研究学者提出了 k-CSP、d-k-CSP、p-RB 等新模型,这些模型通过简单自然的问题实例生成方式,证明了模型具有精确的相变现象。Fan 和 Shen 提出的 k-CSP 模型则通过固定取值域,使约束长度随变量数增长,进一步拓展了 CSP 的理论框架。该模型同样表现出精确的相变现象,并且在相变区域具有指数级的计算复杂性[6] [7]。Zhou 等人提出的 d-k-CSP 模型结合了 RB 模型和 k-CSP 模型的特点,允许取值域和约束长度同时随变量数增长。该模型不仅具有精确的相变现象,还通过有限尺寸缩放分析,进一步揭示了相变区域的宽度随问题规模的增加而缩小[8]。Zhao 等人对于 RB 模型进行了更深入的研究,探讨了一个不同约束组具有不同约束紧度的模型,即 p-RB 模型。该模型是在约束紧度上的更一般的推广,将约束按照一定权重进行划分,同一约束组的约束紧度相等,不同约束组的约束紧度不同,结合理论证明和数据实验论证,阐释了该模型也具有精确的相变现象,具有更广泛的理论价值与现实意义[9]。基于这些新模型的研究,本文在此基础上提出了具有混合约束长度的理论模型  $RB^{\text{mix}}$  模型。相较于 k-CSP 模型固定取值域,使约束长度随变量数增长;d-k-CSP 模型允许取值域和约束长度同时随变量数增长;p-RB 模

型中约束长度相同而约束紧度不同,  $\text{RB}^{\text{mix}}$  模型则通过混合约束长度的设计, 允许约束长度在同一个问题实例中各不相同, 在约束长度的多样性上具有更强的灵活性。能够更好地模拟现实中的复杂约束条件, 旨在进一步拓展 CSP 问题的理论框架。

文献[10]通过提出 RB 模型, 为研究随机 CSP 的相变现象提供了新的工具和理论支持。这篇文章不仅证明了 RB 模型中存在精确的相变点, 还揭示了该模型中存在大量难以解决的实例[11][12]。这些发现为理解随机 CSP 的复杂性提供了重要见解。文献在 2005 年通过统计物理的方法研究了随机 k-SAT 问题中的聚类现象, 并提出了“难 SAT”阶段的概念[13]。文献[14]在 2010 年研究了随机 k-SAT 问题的解空间结构, 证明了解空间在接近可满足性阈值时会分裂成多个独立的聚类。文献[15]和文献[16]在 2015 年和 2022 年分别研究了具有增长域的随机 CSP 模型(如 RB 模型)和种植 CSP 模型的解空间结构。文献[17]通过有限尺寸缩放分析, 定量描述了随机 CSP 的阈值行为, 并证明了相变区域的宽度随问题规模的增加而缩小。他们通过严格的数学方法(如第一矩方法和第二矩方法)证明了解空间在接近可满足性阈值时会发生聚类现象, 并揭示了种植模型中独立相变、分簇相变和孤立相变的存在。这些研究为理解解空间的复杂性提供了重要的理论基础。在这些理论基础背景下, 本文基于混合约束满足问题模型(即  $\text{RB}^{\text{mix}}$  模型), 该模型是在著名的 CSP 和 RB 模型的基础上修正的具有可变取值域和混合约束长度的新模型, 其中取值域是变量数的指数函数, 且所有约束的约束长度各不相同, 来研究  $\text{RB}^{\text{mix}}$  模型的解空间结构和相变现象。在研究解空间结构的理论证明上, 一阶矩、二阶矩为理解解空间复杂性提供了重要工具。

## 2. 预备知识

### 2.1. RB 模型

CSP 问题是由变量集合  $\{x_1, x_2, \dots, x_N\}$ 、约束集合  $\{C_1, C_2, \dots, C_m\}$  以及相容赋值集合  $\{R_{i1}, R_{i2}, \dots, R_{ik}\}$  形成。其中每个变量的可取值域为  $D$ ,  $|D| = N^\alpha$ , ( $\alpha > 0$ ); 任意一个约束, 都是从变量集合  $X$  中随机挑选  $k$  个不同变量满足一定约束条件构成。问题实例是由若干个约束取逻辑且构成。对于实例来说, 求解的实质就是找到一组赋值集合能够满足所有约束条件或无解。生成 RB 模型实例步骤为:

a) 从所有可能的  $C_N^k$  约束中随机、可重复选取  $t = rN \ln N$  个约束构成实例的约束集合  $\{C_i\}_{i=1}^t$ , 其中每个约束所含的变量随机挑选且不重复;

b) 对于实例中的每个约束  $C_i = \{x_{i1}, x_{i2}, \dots, x_{ik}\}$ , 约束中所含变量个数作为其约束范围长度, 从取值域中不重复选择赋值, 不相容赋值集合  $R_i \subseteq D_{i1} \times D_{i2} \times \dots \times D_{ik}$  限制变量的可取值, 其中  $|R_i| = p \cdot d^k$ ,  $0 < p < 1$ ,  $p$  是约束紧度。

这样就可以生成 RB 模型的一个随机实例。

### 2.2. RB 模型相变现象

文献[10]提出了定理 1, 用严格的数学分析方法证明了 RB 模型存在精确的约束密度临界点, 且也可以看出在临界值附件有相变现象的出现。

定理 1 设  $r_s = -\alpha/\ln(1-p)$ , 若  $\alpha > 1/k$ ,  $0 < p < 1$  是两个常数, 且  $k$ ,  $p$  满足不等式  $k \geq 1/(1-p)$ , 则有

$$\lim_{n \rightarrow \infty} \Pr(\text{Sat}) = \begin{cases} 1 & \text{for } r < r_s, \\ 0 & \text{for } r > r_s. \end{cases}$$

其中  $\lim_{n \rightarrow \infty} \Pr(\text{Sat})$  表示问题实例可解的概率。当控制参数较小时, 解空间结构处于复制对称阶段, 这个时候所有解之间的汉明距离相对来说较平均, 分布均匀, 更容易找到解的存在; 随着控制参数变化, 解空

间结构进一步复杂化, 进入复制对称破缺阶段, 此时, 解的分布不再均匀, 所有解会形成独特的聚类, 求解难度显著增大, 难以找到可满足解。

### 2.3. $\text{RB}^{\text{mix}}$ 模型

$\text{RB}^{\text{mix}}$  模型是在  $\text{RB}$  模型基础上进行的扩展, 其主要创新点在于约束长度的混合性。 $\text{RB}^{\text{mix}}$  模型继承了  $\text{RB}$  模型的指数增长取值域和固定约束长度的特点, 但在约束长度的多样性上进行了拓展。 $\text{RB}$  模型中的所有约束具有相同的长度, 而  $\text{RB}^{\text{mix}}$  模型允许约束长度不同。相较于  $\text{RB}$  模型的随机实例是由固定长度的约束组成, 即每个约束都包含个相同数目的变量,  $\text{RB}^{\text{mix}}$  模型则通过构建多组异构约束集合实现约束长度的混合, 也就是将约束集合划分为若干子集, 每个子集中的约束由个不同的变量组成。特别地, 当所有子集约束长度取值相同时,  $\text{RB}^{\text{mix}}$  模型退化为经典的  $\text{RB}$  模型。这种混合约束长度的设计更贴近实际应用中的多样性, 能够更好地模拟复杂问题中的约束结构。具体生成实例步骤如下:

a) 实例中所有的变量  $\{x_i\}_{i=1,\dots,N}$  都有  $d = N^\alpha$  个可能值的取值集合  $D$ ;

b) 将所有约束按照一定比例所含进行分组, 同一组内的约束具有相同的变量个数, 而不同组的约束所含变量个数不同。同一约束组所含变量个数相同, 不同变量组的约束所含变量不同。第  $i$  组约束中随机可重复选择  $c_i r N \ln N$  个约束, 该约束组中每个约束中均涉及  $k_i$  个变量。有  $\sum_{i=1}^l c_i = 1$ ,  $l$  为有限整数。如果  $i \neq j$ , 则  $k_i \neq k_j$ ,  $k_i \geq 2$  是整数;

c) 对于随机选取的每一个约束来说, 有  $Q = pd^{k_i}$  个不相容赋值, 也就是说每个约束都有  $(1-p)d^{k_i}$  个相容赋值集合。

为了方便证明, 引出下列几个定义。

定义 1 赋值对  $\langle t_i, t_j \rangle$

对于随机约束满足问题中任意一个实例  $I$ ,  $t_i = (x_{i1}, x_{i2}, \dots, x_{iN})$ ,  $t_j = (x_{j1}, x_{j2}, \dots, x_{jN})$  是实例的两组赋值, 当且仅当  $t_i$  和  $t_j$  都满足实例时, 赋值对才满足问题实例。

定义 2 Hamming 距离

对于变量集合中相同位置的变量取不同值的个数, 用  $d_H(t_i, t_j)$  表示。从表达式中, Hamming 距离可以简化为

$$d_H(t_i, t_j) = \sum_{a=1}^n 1_{t_{ia} \neq t_{ja}}$$

其中  $1_{t_{ia} \neq t_{ja}}$  是指示函数, 当  $t_{ia} \neq t_{ja}$ ,  $1_{t_{ia} \neq t_{ja}} = 1$ ; 当  $t_{ia} = t_{ja}$ ,  $1_{t_{ia} \neq t_{ja}} = 0$ 。

### 2.4. $\text{RB}^{\text{mix}}$ 模型相关结论

这里我们提出关于  $\text{RB}^{\text{mix}}$  模型存在相变现象的定理。

定理 2 设  $r_s = -\alpha / \ln(1-p)$ ,  $k_0 = \min\{k_1, k_2, \dots, k_l\}$ , 若  $\alpha > 1/k_0$ ,  $0 < p < 1$  为两个常数, 且  $k_0$ ,  $p$  满足不等式  $k_0 \geq 1/(1-p)$ , 则

$$\lim_{n \rightarrow \infty} \Pr(\text{Sat}) = \begin{cases} 1 & \text{for } r < r_s, \\ 0 & \text{for } r > r_s. \end{cases}$$

通过定理分析, 可以知道  $\text{RB}^{\text{mix}}$  模型存在和  $\text{RB}$  模型相同的精确的相变点, 也会表现出相同的相变现象, 从定理内容上也可以看出相变点不会因为约束所含变量个数的不同而变化, 这从理论上严格拓展了  $\text{RB}$  模型只能研究相同约束变量的局限。在控制参数较小时, 模型实例以概率 1 可求解; 随着控制参数变化, 可解概率也出现从 1 到 0 的突然转变, 且求解难度在临界值附近达到峰值。对于  $\text{RB}^{\text{mix}}$  模型, 有着和  $\text{RB}$  模型一样的阈值行为, 那么做出大胆猜测, 解空间结构、相变行为会出现和  $\text{RB}$  模型相同的变化, 在

第三章进行理论证明。

### 3. RB<sup>mix</sup> 模型的分簇现象

本章主要介绍了当 RB<sup>mix</sup> 模型的解空间结构发生变化时，表现出的相变行为。很多 CSP 问题都会出现可解概率从 1 到 0 的突变现象，且在相变点附近求解难度达到峰值。模型的分簇现象、平均计算复杂度均与解空间的高度结构化有关。在标准 RB 模型中，解空间在接近可满足性阈值时会分裂成多个独立的聚类，且每个聚类中的解数量呈指数级增长。RB<sup>mix</sup> 模型的解空间结构表现出同样复杂的特征。通过理论分析，我们发现 RB<sup>mix</sup> 模型在控制参数增加时，解空间会经历从复制对称阶段到复制对称破缺阶段的转变，但并未出现分簇相变。这一现象与标准 RB 模型的结果一致，表明 RB<sup>mix</sup> 模型在解空间结构上具有与 RB 模型相似的阈值行为。

#### 3.1. 可满足性相变现象

可满足相变是 CSP 问题中最先被关注的相变现象。且在相变点附近，问题实例的可解性与不可解性表现出明显的转变。可满足性相变是理解 CSP 复杂性的关键，它揭示了问题从容易解到难以解的转变，这在理论计算机科学中具有重要意义。根据定理 2 可以分析出问题实例发生可满足相变的临界点就是  $r = r_s$  处。在可满足相变之前，问题实例以概率 1 可解；在可满足相变之后，问题实例以概率 0 可解。

#### 3.2. 分簇相变现象

在分簇相变之前，所有的解都位于一个巨大的解簇中，而在分簇相变之后，解空间会分裂成多个独立的聚类，每个聚类中的解数量呈指数级减少。接下来结合矩方法进行证明。

给定 RB<sup>mix</sup> 模型中一个随机实例  $I$ ， $S(I) = \{\sigma; \sigma \text{ 满足 } I\}$  表示实例  $I$  的解空间，设  $\sigma, \tau$  为两组赋值。用  $Z(x)$  定义任意两个解之间 Hamming 距离为  $xN$  的解对数，其中  $t(x)$  表示特定 Hamming 距离条件为  $xN$  下的赋值对数量

$$t(x) = N^{\alpha N} \binom{N}{xN} (N^{\alpha} - 1)^{\alpha N}.$$

对于  $t_i$  和  $t_j$  两组赋值，赋值有两种选择，即赋值对中所有变量都取相同值或赋值对中有变量的值不取同。赋值对是第一种情况概率为  $\frac{C(S, k_i)}{C(N, k_i)}$ ；赋值对是第二种情况概率为  $1 - \frac{C(S, k_i)}{C(N, k_i)}$ 。如果所有变量取

相同赋值，且这样的赋值对是解对数的概率为  $\frac{C(d^{k_i} - 1, pd^{k_i})}{C(d^{k_i}, pd^{k_i})}$ ；除此之外的赋值对是解对数的概率为

$\frac{C(d^{k_i} - 2, pd^{k_i})}{C(d^{k_i}, pd^{k_i})}$ 。用  $q(x)$  表示该 Hamming 距离下的赋值对满足一个约束的概率为

$$q(x) = \frac{C(d^{k_i} - 1, pd^{k_i})}{C(d^{k_i}, pd^{k_i})} \cdot \frac{C(S, k_i)}{C(N, k_i)} + \frac{C(d^{k_i} - 2, pd^{k_i})}{C(d^{k_i}, pd^{k_i})} \cdot \left(1 - \frac{C(S, k_i)}{C(N, k_i)}\right)$$

其中

$$\frac{C(d^{k_i} - 1, pd^{k_i})}{C(d^{k_i}, pd^{k_i})} = \frac{d^{k_i} - q}{d^{k_i}} = 1 - p$$

$$\frac{C(d^{k_i}-2, pd^{k_i})}{C(d^{k_i}, pd^{k_i})} = \frac{(d^{k_i}-1-pd^{k_i})(d^{k_i}-pd^{k_i})}{d^{k_i}(d^{k_i}-1)} = (1-p)^2 + O\left(\frac{1}{d^{k_i}}\right)$$

根据组合公式,  $C(S, k_i) = \frac{S!}{k_i! \cdot (S-k_i)!}$ ,  $C(N, k_i) = \frac{N!}{k_i! \cdot (N-k_i)!}$ , 有

$$\frac{C(S, k_i)}{C(N, k_i)} = \frac{\frac{S}{N}(\frac{S}{N}-\frac{1}{N})(\frac{S}{N}-\frac{2}{N})\cdots(\frac{S}{N}-\frac{k_i-1}{N})}{(1-\frac{1}{N})(1-\frac{2}{N})\cdots(1-\frac{k_i-1}{N})} = (1-x)^{k_i} + \frac{g_i(x)}{N} + O\left(\frac{1}{N^2}\right)$$

当变量规模较大时

$$q(x) = (1-p)\left((1-x)^{k_i} + \frac{g_i(x)}{N}\right) + (1-p)^2\left(1 - (1-x)^{k_i} - \frac{g_i(x)}{N}\right)$$

令  $t = (1-x)^{k_i} + \frac{g_i(x)}{N}$ , 则

$$q(x) = (1-p)t + (1-p)^2(1-t)$$

这样的赋值对满足实例中所有约束的概率是

$$P(1_{\sigma, \tau \in S(1)}) = \prod_{i=1}^l \left[ (1-p)t + (1-p)^2(1-t) \right]^{c_i r N \ln N}$$

设  $Z(x)$  的期望值为  $E(Z(x))$ , 为  $t(x)$  与  $P(1_{\sigma, \tau \in S(1)})$  的乘积

$$E(Z(x)) = \binom{N}{xN} d^N (d-1)^{xN} \prod_{i=1}^l \left[ (1-p)t + (1-p)^2(1-t) \right]^{c_i r N \ln N}$$

由于  $E(Z(x))$  不好计算, 且变量取值域规模是随着变量数呈指数增长。因此, 为了方便计算定义

$$h(x) = \lim_{N \rightarrow \infty} \ln E(Z(x)) / N \ln N$$

根据斯特林公式  $N! \approx \sqrt{2\pi N} \left(\frac{N}{e}\right)^N$ , 可得,

$$\lim_{N \rightarrow \infty} \frac{\ln \binom{N}{xN}}{N} = \ln \left( x^x (1-x)^{1-x} \right);$$

$$\ln d^N = N \ln d;$$

$$\ln (d-1)^{xN} = xN \ln (d-1);$$

$$\begin{aligned} \ln \prod_{i=1}^l \left[ (1-p) \cdot t + (1-p)^2 (1-t) \right]^{c_i r N \ln N} &= \ln \left\{ \left[ (1-p) \cdot t + (1-p)^2 (1-t) \right]^{c_1 r N \ln N} \cdot \left[ (1-p) \cdot t + (1-p)^2 (1-t) \right]^{c_2 r N \ln N} \right. \\ &\quad \left. \cdots \left[ (1-p) \cdot t + (1-p)^2 (1-t) \right]^{c_l r N \ln N} \right\} \\ &= \ln \left[ (1-p) \cdot t + (1-p)^2 (1-t) \right]^{c_1 r N \ln N} + \ln \left[ (1-p) \cdot t + (1-p)^2 (1-t) \right]^{c_2 r N \ln N} \\ &\quad + \cdots + \ln \left[ (1-p) \cdot t + (1-p)^2 (1-t) \right]^{c_l r N \ln N} \\ &= \sum_{i=1}^l c_i r N \ln N \ln \left[ (1-p) \cdot t + (1-p)^2 (1-t) \right] \end{aligned}$$

有



$$h(x) = \lim_{N \rightarrow \infty} \left[ \frac{\ln(x^x (1-x)^{1-x})}{N \ln N} + \frac{N \ln d}{N \ln N} + \frac{xN \ln(d-1)}{N \ln N} + \sum_{i=1}^l c_i r \ln [p(1-p)t + (1-p)^2] \right]$$

$$= \alpha(1+x) + \sum_{i=1}^l c_i r \ln [p(1-p)t + (1-p)^2].$$

当控制参数  $r$  变大时,  $h(x)$  函数是单调递减函数。由 Markov 不等式  $P(Z(x) > 0) \leq E(Z(x))$ , 可以分析得出, 当  $h(x) < 0$  时,  $E(Z(x)) \rightarrow 0$ , 就可以得到  $P(Z(x) = 0) = 1$ , 也就是当  $h(x)$  函数图像在  $x$  轴下方时, 所对应的 Hamming 距离为  $xN$  的赋值对数不存在。对函数进行分析, 可得

$$h'(x) = \alpha + \sum_{i=1}^l c_i r \frac{-pk_i(1-p)(1-x)^{k_i-1}}{p(1-p)(1-x)^{k_i} + (1-p)^2}$$

$$h''(x) = \sum_{i=1}^l c_i r p k_i \frac{(1-x)^{k_i-2} [(1-p)(k_i-1) - p(1-x)^{k_i}]}{[p(1-x)^{k_i} + (1-p)]^2}$$

由定理 2 中  $k_i \geq \frac{1}{1-p}$ , 可得  $h''(x) > 0$ , 即  $h(x)$  是凹函数。令  $r = r_0 r_s$ , 且  $r_s = -\frac{\alpha}{\ln(1-p)}$

- 1) 当  $r = r_s$  时 ( $r_0 = 1$ ),  $h(0) = 0, h(1) = 0$ ;
- 2) 当  $r < r_s$  时 ( $r_0 < 1$ ),  $h(0) > 0, h(1) > 0$ ;
- 3) 当  $r > r_s$  时 ( $r_0 > 1$ ),  $h(0) < 0, h(1) < 0$ 。

接下来, 主要是证明  $h(x)$  函数图像在轴下方时, 这样的赋值对数不存在, 即  $h(x) < 0$ , 赋值对不是解对数

$$P\left(\sum_{xN=x_1N+1}^{x_2N-1} Z(x) > 0\right) \leq E\left(\sum_{xN=x_1N+1}^{x_2N-1} Z(x)\right)$$

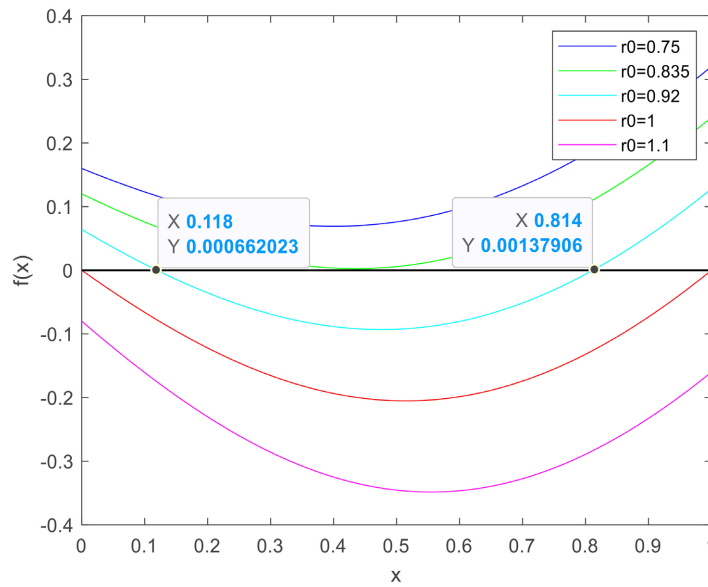
$$\leq (x_2 - x_1) N \cdot \max_{x \in (x_1, x_2)} E(Z(x))$$

$$\rightarrow (x_2 - x_1) N \cdot \max_{x \in (x_1, x_2)} N^{h(x)N}$$

$$\rightarrow 0.$$

这个式子得出不存在 Hamming 距离为  $x_1N$  和  $x_2N$  之间的赋值对数。也就是说解空间所有的解都被分到不同的簇域中, 同一个簇域之间的最大 Hamming 距离为  $x_1N$ , 不同簇域之间最小 Hamming 距离是  $x_2N$ 。分析可得出, 分簇相变出现在可满足相变之后。

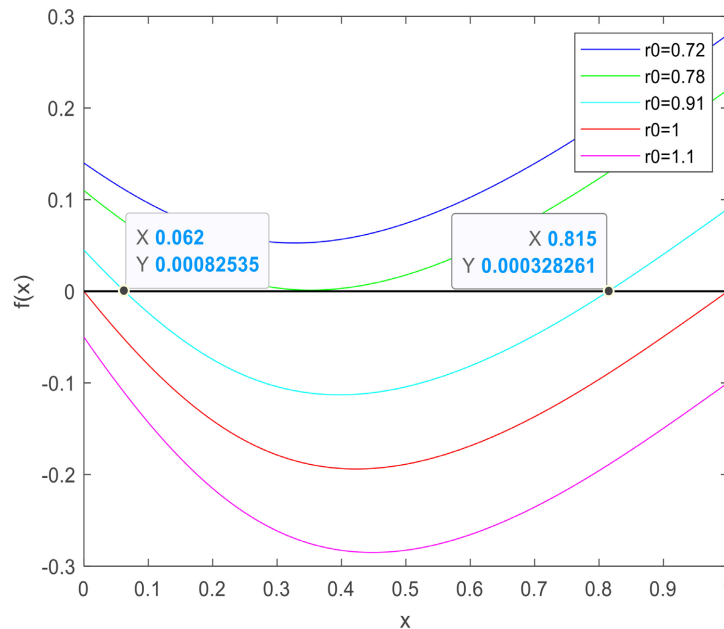
根据上述理论分析, 对  $RB^{\text{mix}}$  模型生成的实例进行赋值, 通过图像直观展示解簇的分布区间。本文探索混合变量个数下的模型的分簇相变, 通过解析实例的解空间结构, 探究解空间何时会出现聚集成簇。在其他参数不变情况下, 根据定理 2 相变点的公式, 我们通过赋三组不同  $k$  值、不同  $\alpha$  值来验证控制参数处于临界值附近时解空间的分簇现象。第一组赋值参数为  $\alpha = 0.8, p = 0.4, K = (2, 3), L = (0.6, 0.4)$ 。根据定理 2 可得  $r_s = 1.5661$ 。设控制参数  $r = r_0 r_s$ , 当  $r_0 < 1$  时, 由定理 2 可知, 在可满足相变点前, 问题实例以概率 1 可解; 当  $r_0 > 1$  时, 在可满足相变点后, 问题实例以概率 0 可解。从图 1 中可以看出, 随着控制参数  $r_0$  变化, 会有相应的函数图像与之对应。举例当  $r_0 = 0.835$  时,  $f(x) = 0$  有一个交点, 当  $0.835 < r_0 < 1$  时,  $f(x) = 0$  有两个交点。通过举  $r_0 < 1$  的一个例子, 即  $r_0 = 0.92$  也就是说此时解空间开始发生分簇。可以发现, 当  $x \in (0.118N, 0.813N)$  时, 函数  $h(x) < 0$ 。也就是不存在 Hamming 距离在  $0.118N$  和  $0.813N$  之间的赋值对数。解空间结构被分解成不同大小的簇域, 其中同一个簇域之间的最大 Hamming 距离为  $0.118N$ , 不同簇域之间最小 Hamming 距离是  $0.813N$ 。



**Figure 1.** Image of  $h(x)$  function with the control parameter of  $\alpha = 0.8, p = 0.4, K = (2, 3), L = (0.6, 0.4)$

**图 1.** 控制参数为  $\alpha = 0.8, p = 0.4, K = (2, 3), L = (0.6, 0.4)$  的  $h(x)$  函数图像

第二组赋值参数为  $\alpha = 0.5, p = 0.6, K = (3, 6), L = (0.6, 0.4)$ 。图 2 中曲线从上到下  $r$  值分别为  $0.75r_s, 0.835r_s, 0.92r_s, r_s, 1.1r_s$ 。图中  $x \in (0.78r_s, r_s)$  区域内，函数与坐标轴有两个交点。用  $0.91r$  举例，可以发现当  $x \in (0.062, 0.815)$ ，函数  $h(x) < 0$ 。也就是不存在 Hamming 距离在 Hamming 和  $0.815N$  之间的赋值对数。解空间结构被分解成不同大小的簇域，其中同一个簇域之间的最大 Hamming 距离为  $0.062N$ ，不同簇域之间最小 Hamming 距离是  $0.815N$ 。

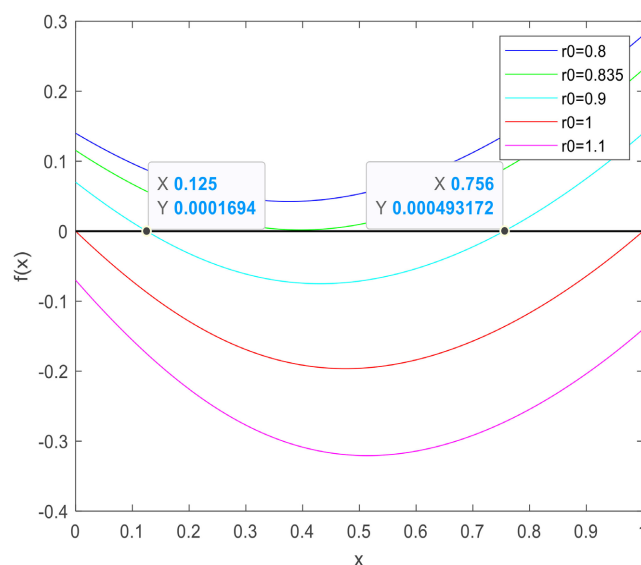


**Figure 2.** Image of  $h(x)$  function with the control parameter of  $\alpha = 0.5, p = 0.6, K = (3, 6), L = (0.6, 0.4)$

**图 2.** 控制参数为  $\alpha = 0.5, p = 0.6, K = (3, 6), L = (0.6, 0.4)$  的  $h(x)$  函数图像



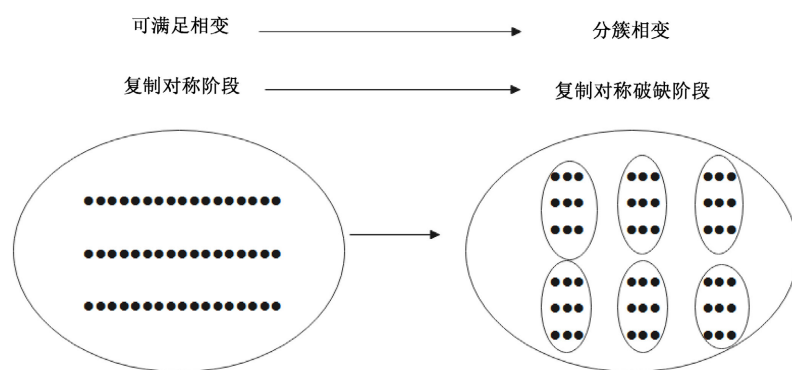
第三组赋值参数分别为  $\alpha = 0.75, p = 0.45, K = (2, 4), L(0.6, 0.4)$ 。举例  $r_0 = 0.9$ ，从图 3 中可以发现当  $x \in (0.125, 0.756)$ ，有函数  $h(x) < 0$ ，也就是说不存在 Hamming 距离为  $0.125N$  和  $0.756N$  之间的赋值对数。解空间被分解成不同大小的簇域，其中同一个簇域的最大 Hamming 距离为  $0.125N$ ，不同簇域的最小 Hamming 距离为  $0.756N$ 。



**Figure 3.** Image of  $h(x)$  function with the control parameter of  $\alpha = 0.75, p = 0.45, K = (2, 4), L(0.6, 0.4)$

**图 3.** 控制参数为  $\alpha = 0.75, p = 0.45, K = (2, 4), L(0.6, 0.4)$  的  $h(x)$  函数图像

具体而言， $\text{RB}^{\text{mix}}$  模型的解空间在控制参数较低时处于复制对称阶段，解分布均匀，算法求解较为容易。随着控制参数的增加，解空间进入复制对称破缺阶段，解分布不再均匀，解空间分裂成多个聚类，每个聚类中的解数量呈指数级减少。然而，与  $k\text{-SAT}$  问题不同， $\text{RB}^{\text{mix}}$  模型并未出现凝聚相变，即不存在一个有限的聚类包含几乎所有解的情况，进一步验证了  $\text{RB}^{\text{mix}}$  模型在解空间结构上的独特性。



此外， $\text{RB}^{\text{mix}}$  模型的解空间结构还表现出与标准  $\text{RB}$  模型相似的分簇现象。在分簇相变之前，所有解都位于一个巨大的解簇中；而在分簇相变之后，解空间分裂成多个独立的聚类，每个聚类中的解数量呈指数级减少。这一现象通过理论分析和数值实验得到了验证，进一步支持了  $\text{RB}^{\text{mix}}$  模型在解空间结构上的复杂性。

综上所述,  $\text{RB}^{\text{mix}}$  模型在控制参数增加时表现出与标准  $\text{RB}$  模型相似的分相行为, 但并未出现凝聚相变。这一发现为理解随机 CSP 的解空间复杂性提供了新的视角, 并为设计高效算法提供了理论支持。未来的研究可以进一步探讨  $\text{RB}^{\text{mix}}$  模型在更高控制参数下的解空间结构, 以及其在实际应用中的表现。

#### 4. 总结

解空间结构的研究是随机约束满足问题领域的前沿课题, 本文基于  $\text{RB}$  模型提出了  $\text{RB}^{\text{mix}}$  模型, 通过引入混合约束长度, 进一步揭示了高度结构化的解空间特征。关于各种模型的解空间结构既有客观统一性, 又存在彼此的差异性, 这为我们关于约束满足问题提供了更全面的理论支持, 帮助更好地理解计算复杂性。未来的研究可以进一步深化解空间结构的理论研究, 尝试通过更复杂的数学工具(如高阶矩方法)来探索解空间的复杂性, 算法上可以结合更多优化算法解决更大规模的随机 CSP 问题。根据解空间结构特性设计高效算法的核心在于理解问题的内在复杂性和解的分布规律, 通过分析解空间的聚集特性、相变点附近的特性可以为算法设计提供重要线索。通过分析解空间的结构变化, 可以设计出针对不同阶段的算法策略, 从而提高求解效率。例如, 在控制参数较小时, 解空间结构简单, 问题实例几乎总是可解的。此时, 算法设计可以侧重于快速找到可行解, 可以选择约束最多的变量优先赋值; 控制参数较大时, 解空间结构复杂, 解空间分裂成多个独立的聚类, 问题实例可能不可解。此时, 算法设计需要应对解空间的复杂性和求解难度, 可以选择从多个初始解出发, 探索解空间的不同区域, 避免陷入局部最优; 在相变点附近, 问题实例的求解难度达到最大值, 解空间结构发生显著变化。此时, 算法设计需要应对解空间的不稳定性和求解难度的急剧增加, 可以利用机器学习模型预测解空间结构的变化, 指导算法设计。

#### 参考文献

- [1] 杜立智, 陈和平, 符海东. NP 完全问题研究及前景剖析[J]. 武汉工程大学学报, 2015, 37(10): 73-78.
- [2] Smith, B.M. and Dyer, M.E. (1996) Locating the Phase Transition in Binary Constraint Satisfaction Problems. *Artificial Intelligence*, **81**, 155-181. [https://doi.org/10.1016/0004-3702\(95\)00052-6](https://doi.org/10.1016/0004-3702(95)00052-6)
- [3] Gent, I.P., Macintyre, E., Prosser, P., Smith, B.M. and Walsh, T. (2001) Random Constraint Satisfaction: Flaws and Structure. *Constraints*, **6**, 345-372. <https://doi.org/10.1023/a:1011454308633>
- [4] Prosser, P. (1996) An Empirical Study of Phase Transitions in Binary Constraint Satisfaction Problems. *Artificial Intelligence*, **81**, 81-109. [https://doi.org/10.1016/0004-3702\(95\)00048-8](https://doi.org/10.1016/0004-3702(95)00048-8)
- [5] Achlioptas, D., Molloy, M.S.O., Kirousis, L.M., Stamatiou, Y.C., Kranakis, E. and Krizanc, D. (2001) Random Constraint Satisfaction: A More Accurate Picture. *Constraints*, **6**, 329-344. <https://doi.org/10.1023/a:1011402324562>
- [6] Fan, Y. and Shen, J. (2011) On the Phase Transitions of Random K-Constraint Satisfaction Problems. *Artificial Intelligence*, **175**, 914-927. <https://doi.org/10.1016/j.artint.2010.11.004>
- [7] Zhou, G., Gao, Z. and Liu, J. (2014) On the Constraint Length of Random k-CSP. *Journal of Combinatorial Optimization*, **30**, 188-200. <https://doi.org/10.1007/s10878-014-9731-3>
- [8] Zhou, G., Gao, Z. and Liu, J. (2016) The Scaling Window of the Model d-k-CSP. *Journal of Mathematical Analysis and Applications*, **434**, 342-352. <https://doi.org/10.1016/j.jmaa.2015.09.015>
- [9] 赵春艳, 范如梦, 刘雅楠. 不同紧度下约束满足问题的相变现象[J]. 计算机应用研究, 2020, 37(9): 2739-2743.
- [10] Xu, K. and Li, W. (2000) Exact Phase Transitions in Random Constraint Satisfaction Problems. *Journal of Artificial Intelligence Research*, **12**, 93-103. <https://doi.org/10.1613/jair.696>
- [11] Xu, K. and Li, W. (2006) Many Hard Examples in Exact Phase Transitions. *Theoretical Computer Science*, **355**, 291-302. <https://doi.org/10.1016/j.tcs.2006.01.001>
- [12] Šulc, P. and Zdeborová, L. (2010) Belief Propagation for Graph Partitioning. *Journal of Physics A: Mathematical and Theoretical*, **43**, Article ID: 285003. <https://doi.org/10.1088/1751-8113/43/28/285003>
- [13] Mézard, M., Mora, T. and Zecchina, R. (2005) Clustering of Solutions in the Random Satisfiability Problem. *Physical Review Letters*, **94**, Article ID: 197205. <https://doi.org/10.1103/physrevlett.94.197205>
- [14] Achlioptas, D., Coja-Oghlan, A. and Ricci-Tersenghi, F. (2011) On the Solution-Space Geometry of Random Constraint Satisfaction Problems. *Random Structures & Algorithms*, **38**, 251-268. <https://doi.org/10.1002/rsa.20323>

- 
- [15] Xu, W., Zhang, P., Liu, T. and Gong, F. (2015) The Solution Space Structure of Random Constraint Satisfaction Problems with Growing Domains. *Journal of Statistical Mechanics: Theory and Experiment*, **2015**, P12006. <https://doi.org/10.1088/1742-5468/2015/12/p12006>
  - [16] Xu, W. and Zhang, Z. (2022) The Solution Space Structure of Planted Constraint Satisfaction Problems with Growing Domains. *Journal of Statistical Mechanics: Theory and Experiment*, **2022**, Article ID: 033401. <https://doi.org/10.1088/1742-5468/ac4e86>
  - [17] Zhao, C. and Zheng, Z. (2011) Threshold Behaviors of a Random Constraint Satisfaction Problem with Exact Phase Transitions. *Information Processing Letters*, **111**, 985-988. <https://doi.org/10.1016/j.ipl.2011.07.006>