Автоматика и телемеханика, № 1, 2020

© 2020 г. С.И. УВАРОВ, канд. техн. наук (uvarov53@gmail.com)
(Институт проблем управления им. В.А. Трапезникова РАН, Москва)

УСОВЕРШЕНСТВОВАННЫЙ ГЕНЕРАТОР 3-КНФ ФОРМУЛ

Рассматривается задача выполнимости (SAT-problem) булевых формул, заданных в конъюнктивной нормальной форме с ограничением, что каждый дизъюнкт содержит по три литерала переменных (3-КНФ). В эмпирических исследованиях широко используется генерация случайных формул с фиксированной длиной дизъюнкта. Феноменом этого метода является многократно подтвержденная линейная зависимость числа дизъюнктов формулы от числа булевых переменных в точке «фазового перехода» – от статуса выполнимых к статусу невыполнимых (когда доля невыполнимых формул становится превалирующей). Предложен и исследован метод генерации случайных формул, имеющий меньший коэффициент (3,49) пропорциональности между числом дизъюнктов и числом переменных в точке «фазового перехода» (для известного метода генерации этот коэффициент равен 4,23).

Ключевые слова: задача выполнимости (SAT-problem), конъюнктивная нормальная форма (КНФ), дизъюнкт, литерал, булевы переменные.

DOI: 10.31857/S0005231020010110

1. Введение

В работе исследуется генерация сложных примеров для задачи выполнимости логических формул (SAT-problem). Задача выполнимости является опорной для обширного класса NP-complete. Многие практически важные задачи несложными преобразованиями могут быть приведены к задаче выполнимости. Таковыми, например, являются задачи верификации аппаратуры и программного обеспечения, планирования, составления расписаний, комбинаторного анализа [1]. Важнейшим применением задачи выполнимости является автоматическое доказательство теорем – основа искусственного интеллекта.

Задача выполнимости относится к труднорешаемым задачам, при этом вопрос о возможности существования решения, алгоритмическая сложность которого ограничена некоторым полиномом от длины записи логической формулы, остается открытым.

Многие из встречающихся частных примеров в действительности оказываются достаточно простыми. Как правило, достаточно простыми являются и взятые из практики (индустриальные) примеры.

Поиск путей построения сложных примеров важен как для понимания природы сложности задачи, так и для построения тестовых примеров (benchmarks) с целью экспериментальной оценки алгоритмов [2]. Сложные примеры активно способствуют выявлению недостатков в разработанных алгоритмах и тем самым указывают пути их дальнейшего совершенствования.

Считается, что относительно сложные примеры удается строить с помощью случайных чисел. Генерируемые при помощи случайных чисел формулы широко изучаются, поскольку это достаточно естественная модель, которая проливает свет на фундаментальные структурные свойства задачи выполнимости. Такие формулы хорошо отражают специфику задач в системах логических доказательств [3].

Первое известное значимое исследование было выполнено на формулах со случайной длиной дизъюнктов. Было продемонстрировано [4], что при некоторых параметрах генератора таких формул задача выполнимости для них может быть решена в среднем за $O(MN^2)$ шагов. Последующие исследования [2–6] показали, что семейство формул, со случайной длиной дизъюнктов, должно рассматриваться как простое. Простота анализа формул со случайной длиной дизъюнктов объясняется тем, что они часто содержат пустые дизъюнкты, дизъюнкты из единственного литерала (юниты) и тривиальные дизъюнкты. В настоящее время эмпирические исследования смещены к методам построения формул, в которых генерация таких простых дизъюнктов исключена [7, 8]. Подробное изложение результатов по рассматриваемой тематике содержится в [1].

Подавляющее большинство исследователей используют вариации генератора с фиксированной длиной дизъюнкта. Для 3-КНФ длина дизъюнкта равна трем. Литералы переменных, включенные в дизъюнкт, выбираются с использованием последовательности случайных чисел. Такой генератор хорош тем, что если подать на его вход нужную последовательность чисел, он способен породить любую формулу. Он очень удобен для теоретических исследований: его простота позволила аналитически доказать, что (в пределе) при увеличении количества переменных при $R \leqslant 3{,}52$ (R – отношение числа дизъюнктов к числу переменных) генератор порождает пренебрежимо малое число невыполнимых формул, а при $4{,}51 \leqslant R$ алгоритм порождает пренебрежимо малое число выполнимых формул [9, 10].

Известна интересная гипотеза [5] о том, что нетривиальные формулы, являющиеся невыполнимыми при меньшем числе ограничений (дизъюнктов), как правило являются более сложными для доказательства невыполнимости – анализа. Поэтому для построения сложных примеров представляет интерес поиск способов генерации формул, значительный процент которых является невыполнимыми при $R \leqslant 3{,}52$.

2. Генераторы случайных формул

Будем использовать следующие обозначения. $\mathbf{X}$ – множество из N булевых переменных $X_\imath$, $\mathbf{X} = \{X_1, \ldots, X_N\}$. Литералами переменной $X_\imath$ назовем термы x_i^0 и x_i^1, где $x_i^0 = X_i$ и $x_i^1 = \neg X_i$.

Рассматриваются формулы $F(M) = C_1^3 \wedge C_2^3 \wedge \ldots \wedge C_M^3$, представленные в конъюнктивной нормальной форме 3-КНФ, являющиеся конъюнкцией набора из M трехлитеральных дизъюнктов C_j^3. Исследуется их выполнимость, т.е.

возможность отыскания набора значений булевых переменных, на которых формула принимает значение «истина».

Трехлитеральным дизъюнктом $C_j^3(x_k^p, x_\ell^q, x_m^r)$ является дизъюнкция трех литералов переменных $X_\imath$ из множества $\mathbf{X}$. Пример дизъюнкта с конкретными значениями верхних индексов литералов: $C_j^3(x_k^0, x_\ell^1, x_m^0) = x_k^0 \vee x_\ell^1 \vee x_m^0 = X_k \vee \neg X_\ell \vee X_m$.

Большое число публикаций, связанных с задачей выполнимости булевых формул, заданных в 3-КНФ, представляют результаты эмпирических исследований. Основным инструментом таких исследований является генератор случайных формул. Обычно для заданного числа N переменных желательно строить набор из V невыполнимых формул F_v $(v = 1, \ldots, V)$.

Формула, не имеющая дизъюнктов, выполнима. Искомая невыполнимая формула строится так, что, пока она остается выполнимой, к ранее построенным дизъюнктам добавляется новый. При получении невыполнимого набора дизъюнктов процесс генерации формулы заканчивается. После того, как формула стала невыполнимой, добавление новых дизъюнктов не может сделать ее выполнимой.

Формулу F_v удобно продуцировать на основе потока дизъюнктов. Принадлежащие такому потоку дизъюнкты строятся на основе одной последовательности $\mathbf{S}_v$ случайных натуральных чисел ξ_ν, предположительно равномерно распределенных в диапазоне от 1 до $2N(N-1)(N-2)$.

Разные формулы строятся на основе отличающихся друг от друга последовательностей случайных чисел.

Первый генератор с независимым выбором литералов (далее НВЛ-генератор).

Для построения очередного дизъюнкта C_{j+1} формулы F_v (в предположении, что при построении предшествующих дизъюнктов использованы ω чисел из $\mathbf{S}_v$) выбираются три последовательных случайных числа $\xi_{\omega+1}$, $\xi_{\omega+2}$, $\xi_{\omega+3}$. В произвольным образом упорядоченном множестве $\mathbf{A}$, состоящем из $2N$ литералов от N, переменных выбираем элементы, имеющие номера $\xi_{\omega+1} \pmod{2N}$, $\xi_{\omega+2} \pmod{2N}$ и $\xi_{\omega+3} \pmod{2N}$. Пусть этими элементами являются литералы x_k^p, x_ℓ^q и x_m^r, из них строим очередной дизъюнкт $C_{j+1}(x_k^p, x_\ell^q, x_m^r)$.

После построения невыполнимого набора дизъюнктов процесс генерации формулы F_v заканчивается.

При использовании такого генератора с небольшой вероятностью (порядка $1/N^2$) продуцируется дизъюнкт, содержащий более одного литерала одной переменной. В этом случае дизъюнкт будет тавтологией, если верхние индексы литералов, представляющих одну переменную, различны, либо будет дизъюнктом, в котором число различных литералов меньше трех. Такая ситуация уменьшает сложность анализа формулы, поэтому следует предпочесть следующий генератор.

Второй генератор формул без простых дизъюнктов (далее БПД-генератор) исключает возможность построения дизъюнкта, содержащего более одного литерала одной переменной.

Для построения очередного дизъюнкта C^3_{j+1} формулы F_v из $\mathbf{S}_v$ выбираются три последовательных случайных числа $\xi_{\omega+1}, \xi_{\omega+2}, \xi_{\omega+3}$. Затем в упорядоченном множестве $\mathbf{A}$ выбирается элемент x^p_k с номером $\xi_{\omega+1} \pmod{2N}$. Этот элемент полагается первым литералом дизъюнкта. Второй литерал назначается из множества $\mathbf{A} - \{x^0_k, x^1_k\}$, содержащего $2N-2$ элементов. Выбирается элемент x^q_ℓ с номером $\xi_{\omega+2} \pmod{2N-2}$. Третий литерал дизъюнкта выбирается из множества $\mathbf{A} - \{x^0_k, x^1_k, x^0_\ell, x^1_\ell\}$, содержащего $2N-4$ элементов. Это элемент с номером $\xi_{\omega+2} \pmod{2N-4}$. Выбранный элемент x^r_m становится третьим литералом дизъюнкта $C^3_{j+1}(x^p_k, x^q_\ell, x^r_m)$.

После построения невыполнимого набора дизъюнктов процесс генерации формулы F_v заканчивается.

Поскольку при больших N вероятность того, что дизъюнкт формулы, построенной первым генератором, содержит литералы одной переменной, невелика (порядка $1/N^2$), то результаты эмпирических исследований с этими генераторами близки.

Отметим, что существенная разница в частоте использования различных переменных обычно является одним их факторов упрощения анализа формул и часто используется алгоритмами решения задачи выполнимости (SAT-solver). С целью получения более сложных для анализа формул в настоящей публикации предлагается третий генератор, особенностью которого является выравнивание частоты использования в формуле литералов различных переменных.

Третий генератор, выравнивающий нагрузку на литералы (далее ВНЛ-генератор), использует статистическую информацию о количестве использования литералов переменных в построенных дизъюнктах.

Для построения очередного дизъюнкта C^3_{j+1} формулы F_v выбираем литерал, который минимальное число T раз встречается в ранее построенных дизъюнктах. Пусть это литерал x^t_f переменной X_f. Его полагаем первым литералом генерируемого дизъюнкта.

В отличие от первых двух, третий генератор при задании первого литерала очередного дизъюнкта не использует элементы последовательности $\mathbf{S}_v$ случайных чисел.

Для выбора второго и третьего литералов дизъюнкта используются числа $\xi_{\eta+1}$ и $\xi_{\eta+2}$ из $\mathbf{S}_v$ (η чисел использованы при построении предшествующих дизъюнктов формулы F_v).

Для генерации второго литерала дизъюнкта строится множество $\mathbf{B}$ литералов (переменных, отличных от переменной X_f), входящих в построенные дизъюнкты формулы не более чем $T+2$ раз. Если множество оказалось пустым, повторяем построение множества $\mathbf{B}$ с новым увеличенным на единицу значением T ($T = T+1$). Если множество не пусто, фиксируем число его элементов B и порядок элементов в множестве $\mathbf{B}$. Вторым литералом назначаем элемент x^q_ℓ, имеющий номер $\xi_{\eta+1} \pmod{B}$ в множестве $\mathbf{B}$.

При построении третьего литерала дизъюнкта снова строится множество $\mathbf{B}$ литералов (переменных, отличных от переменных X_f и X_ℓ), входящих в построенные дизъюнкты формулы не более чем $T+2$ раз. Если множество оказалось пустым, повторяем построение множества $\mathbf{B}$ с новым увеличенным

Распределение литералов в невыполнимой формуле, постоенной вторым генератором

19					1											
18					1						1					
17					1					1	1					
16					1					1	1					
15				1	1					1	1		1			
14	1			1	1					1	1		1	1		
13	1			1	1			1		1	1	1	1	1		1
12	1			1	1			1		1	1	1	1	1		1
11	1			1	1			1		0	0	0	1	1		1
10	0	1	1	1	0			1		0	0	0	1	1		1
9	0	1	1	1	0	1	1	1		0	0	0	0	1		1
8	0	1	1	0	0	1	1	1		0	0	0	0	0	1	1
7	0	1	1	0	0	1	1	1	1	0	0	0	0	0	1	1
6	0	1	1	0	0	0	0	0	1	0	0	0	0	0	1	1
5	0	0	0	0	0	0	0	0	1	0	0	0	0	0	1	1
4	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0
3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	X_1	X_2	X_3	X_4	X_5	X_6	X_7	X_8	X_9	X_{10}	X_{11}	X_{12}	X_{13}	X_{14}	X_{15}	X_{16}
1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
2	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
3	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
4	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
5	1	1	1	1	1	1	1	1	1	1	1	0	1	1	1	0
6	1	0	0	1	1	1	0	1	0	0	0	0	1	0	1	0
7	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
8	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
9	0	0	0	0	0	0	0	0	0	0	0		0	0	0	0
10	0	0	0	0	0	0	0	0	0	0	0		0		0	0
11	0			0	0	0	0	0					0		0	
12					0								0		0	

Распределение литералов в невыполнимой формуле, постоенной третьим генератором

Рис. 1. Иллюстрация неравномерности использования литералов в 3-КНФ формулах, построенных БПД- и ВНЛ-генераторами.

на единицу значением T ($T = T + 1$). Если множество не пусто, фиксируем число его элементов B и порядок элементов в множестве $\mathbf{B}$. Третьим литералом назначаем элемент x_m^r, имеющий номер $\xi_{\eta+2} \pmod B$ в множестве $\mathbf{B}$. Построение дизъюнкта $C_{j+1}^3(x_f^t, x_\ell^q, x_m^r)$ завершено.

После построения невыполнимого набора дизъюнктов процесс генерации формулы F_v заканчивается.

На рис. 1 представлены две гистограммы, иллюстрирующие количественное присутствие литералов 16 переменных в двух невыполнимых формулах, построенных ВНЛ-генератором (внизу) и БПД-генератором (вверху). Литералы, представляющие переменные без инверсии, обозначены нулем по значению верхнего индекса, а литералы, представляющие инвертированные переменные, обозначены единицей. В формуле от 16 переменных, построенной третьим генератором, число использованных литералов каждого типа отличается не более чем на два. В формуле от 16 переменных, построенной вторым генератором, 15-я переменная без инверсии используется трижды, а, например, 10-я переменная – одиннадцать раз.

Отметим, что все три описанных генератора могут породить формулу, дизъюнкты которой имеют одинаковые наборы литералов. Вероятность того, что в формуле из $M = kN$ дизъюнктов два будут содержать одинаковые наборы литералов, будет порядка $1/N^2$. Присутствие в формуле таких дизъюнктов не приводит к существенному упрощению ее анализа. Поэтому усложнение генераторов, направленное на исключение возможности генерации формул с дизъюнктами, содержащими идентичные наборы литералов, считается нецелесообразным.

3. Эмпирическое исследование генератора

Замечательным свойством второго БПД-генератора является соблюдаемая с высокой точностью линейная зависимость $\mathfrak{M}_2$ от N.

Если генерировать формулы содержащие по $\lfloor \mathfrak{M}(N) \rfloor$ дизъюнктов от N переменных, математическое ожидание вероятности выполнимости формул будет больше либо равно 0,5. Если генерировать формулы содержащие по $\lceil \mathfrak{M}(N) \rceil$ дизъюнктов от N переменных, математическим ожиданием вероятности невыполнимости будет величина большая, либо равная 0,5.

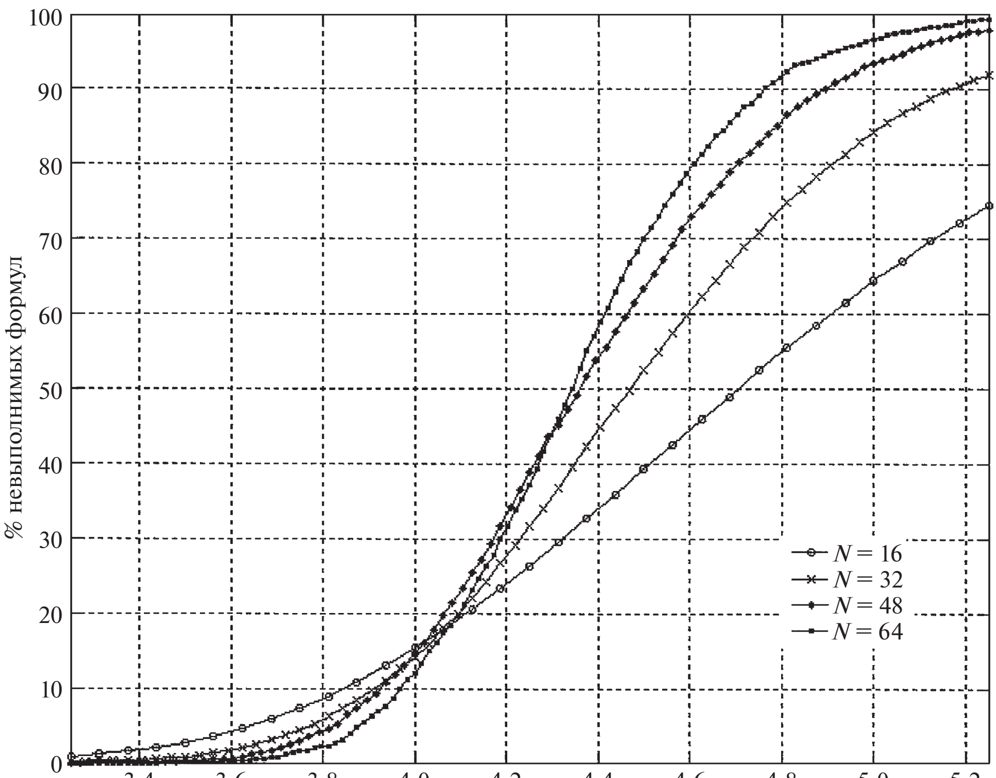

Рис. 2. Иллюстрация статистики перехода от статуса выполнимости к статусу невыполнимости для 3-КНФ формул, порожденных БПД-генератором.

Этот удивительный результат [6] до сих пор не нашел удовлетворительного объяснения.

Феномен линейной зависимости от N точки «фазового перехода» 3-КНФ формул (порожденных БПД-генератором) от статуса выполнимых к статусу невыполнимых был многократно подтвержден последующими публикациями, например [7, 8], и его статистическая достоверность не вызывает сомнения.

Для большей наглядности график процентной доли формул, ставших невыполнимыми после включения в их состав M дизъюнктов, строят относительно величины $R = M/N$, являющейся отношением числа дизъюнктов к числу переменных. На рис. 2 совместно представлены такие зависимости для формул с числом переменных 16, 32, 48, 64, построенных БПД-генератором. Результированы данные по 2000, 4000, 16000 и 32000 формулам для $N = 64$, $N = 48$, $N = 32$ и $N = 16$. В выбранных координатах на графике «фазовый переход» становится все более крутым с увеличением числа переменных.

Проведенные исследования [6–8] показали колоколообразную зависимость медианной сложности анализа формул от числа дизъюнктов. При этом максимальная медианная сложность имела место в непосредственной близости от точки «фазового перехода».

Медианная сложность в определенном смысле является усредненной сложностью анализа формул, содержащих фиксированное число дизъюнктов. В одной из пионерских работ [5], ориентированных на построение сложных для анализа формул, отмечено, что самыми сложными, как правило, оказываются формулы, являющиеся невыполнимыми при минимальном числе дизъюнктов.

Этот эмпирический результат объяснен [5] тем, что задача выполнимости, как и большинство других NP-comlete задач, может рассматриваться как поиск решения, удовлетворяющего определенным ограничениям. Интуитивно ясно, что если ограничений мало, то решение найти легко, так как при этом обычно имеет место множество возможных решений. Аналогично, если ограничений слишком много, достаточно интеллектуальный алгоритм обычно способен быстро отбрасывать большинство тупиковых ветвлений в дереве поиска. Таким образом, разумно ожидать, что самые сложные задачи – это задачи, которые, с одной стороны, не перегружены ограничениями, с другой стороны, имеющиеся ограничения оставляют возможность лишь для небольшого числа решений [6]. Эмпирическими исследованиями подтверждено, что это действительно имеет место [6–8].

Приведенные рассуждения являются мотивацией для поиска алгоритмов построения 3-КНФ формул, являющихся невыполнимыми при меньшем числе дизъюнктов, чем в невыполнимых формулах, построенных ставшими классическими НВЛ- и БПД-генераторами.

На рис. 3 представлены результаты исследований, для формул с числом переменных 16, 32, 48, 64, построенных ВНЛ-генератором. Для $N = 64$, $N = 48$, $N = 32$ и $N = 16$ построено 2000, 4000, 16000 и 32000 формул соответственно.

Как и для второго БПД-генератора, «фазовый переход» на графике рис. 3 становится все более крутым с увеличением числа переменных. Заметно,

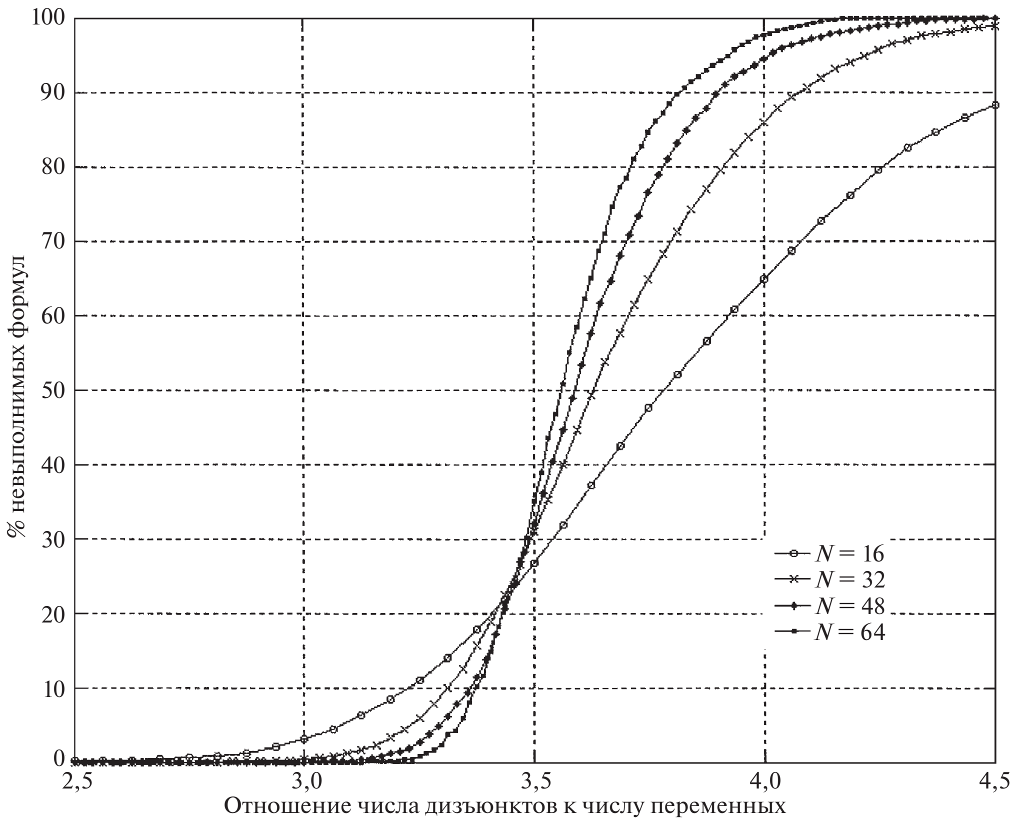

Рис. 3. Иллюстрация статистики перехода от статуса выполнимости к статусу невыполнимости для 3-КНФ формул, порожденных ВНЛ-генератором.

что при одинаковом числе переменных для формул, порожденных третьим ВНЛ-генератором, «фазовый переход» более крутой, чем для формул, порожденных БПД-генератором (см. рис. 2). Важно, что при заданном числе переменных «фазовый переход» для формул, построенных ВНЛ-генератором,

Таблица 1. Связь эмпирических вероятностей невыполнимости формулы со значениями $\lfloor \mathfrak{M}(N) \rfloor$ и $\lceil \mathfrak{M}(N) \rceil$

Второй генератор	Третий генератор
$\lfloor \mathfrak{M}_2(3) \rfloor = 19$ (47,835) $\lceil \mathfrak{M}_2(3) \rceil = 20$ (53,056)	$\lfloor \mathfrak{M}_3(3) \rfloor = 14$ (44,160) $\lceil \mathfrak{M}_3(3) \rceil = 15$ (51,900)
$\lfloor \mathfrak{M}_2(16) \rfloor = 75$ (48,919) $\lceil \mathfrak{M}_2(16) \rceil = 76$ (52,366)	$\lfloor \mathfrak{M}_3(16) \rfloor = 60$ (47,503) $\lceil \mathfrak{M}_3(16) \rceil = 61$ (52,031)
$\lfloor \mathfrak{M}_2(32) \rfloor = 143$ (48,756) $\lceil \mathfrak{M}_2(32) \rceil = 144$ (52,419)	$\lfloor \mathfrak{M}_3(32) \rfloor = 116$ (47,503) $\lceil \mathfrak{M}_3(32) \rceil = 117$ (53,656)
$\lfloor \mathfrak{M}_2(48) \rfloor = 209$ (49,000) $\lceil \mathfrak{M}_2(48) \rceil = 210$ (51,500)	$\lfloor \mathfrak{M}_3(48) \rfloor = 172$ (48,825) $\lceil \mathfrak{M}_3(48) \rceil = 173$ (53,350)
$\lfloor \mathfrak{M}_2(64) \rfloor = 278$ (49,850) $\lceil \mathfrak{M}_2(64) \rceil = 279$ (52,700)	$\lfloor \mathfrak{M}_3(64) \rfloor = 227$ (46,800) $\lceil \mathfrak{M}_3(64) \rceil = 228$ (50,700)

происходит при существенно меньшем соотношении числа дизъюнктов к числу переменных, чем для формул, построенных БПД-генератором.

Процент невыполнимых формул построенных БПД- и ВНЛ-генераторами при заданном числе переменных и заданном числе дизъюнктов, представлен в табл. 1.

При $N = 3$ вероятность того, что содержащая M дизъюнктов формула, порожденная БПД-генератором, будет невыполнимой, может быть вычислена аналитически. Если переменных всего три, то для получения невыполнимой формулы необходимо и достаточно, чтобы среди дизъюнктов формулы присутствовали восемь различных дизъюнктов, каждый из которых содержит литералы различных переменных. Если общее число дизъюнктов в формуле меньше восьми, формула выполнима.

Вероятность того, что формула, содержащая восемь случайных дизъюнктов, будет невыполнимой, равна $7!/8^7 = 0{,}0024$. При увеличении числа дизъюнктов в формуле вероятность того, что формула невыполнима, будет возрастать, стремясь к единице.

Как указано в [6], вероятность генерирования невыполнимой формулы, составленной из M дизъюнктов, соответствует вероятности того, что за M обращений к генератору случайных чисел с диапазоном (1:8) будут сгенерированы все восемь возможных чисел (т.е. в полученной последовательности будет представлено каждое из восьми чисел).

Обозначим через P_M^k вероятность того, что среди M дизъюнктов k будут различными. Вероятности P_M^k $(1 \leqslant k \leqslant 8)$ связаны рекуррентными соотношениями:

$$P_M^k = (k/8)P_{M-1}^k + ((9-k)/8)P_{M-1}^{k-1}$$

при начальных условиях $P_0^0 = 1, P_0^j = 0, j = \overline{1,8}$ и граничных условиях $P_M^0 = 0$.

Вычисления по этим формулам дают $P_{19}^8 = 0{,}478348$, а $P_{20}^8 = 0{,}530558$. Таким образом, для трех переменных точка «фазового перехода» находится между 19 и 20 дизъюнктами.

В ячейках табл. 1 в круглых скобках приведены процентные доли невыполнимых формул.

Аппроксимация методом наименьших квадратов данных, представленных в табл. 1, подтверждает хорошо известную для БПД-генератора линейную зависимость числа $\mathfrak{M}_2$ от числа переменных N.

$$\mathfrak{M}_2(N) = 4{,}23N + 7{,}18.$$

Аппроксимация представленных в табл. 1 эмпирических данных выявляет линейную зависимость $\mathfrak{M}_3$ от числа переменных N и для ВНЛ-генератора.

$$\mathfrak{M}_3(N) = 3{,}49N + 4{,}46.$$

Обе эти зависимости представлены на рис. 4. Полученные зависимости позволяют предположить, что величины 4,23 и 3,49, возможно, являются

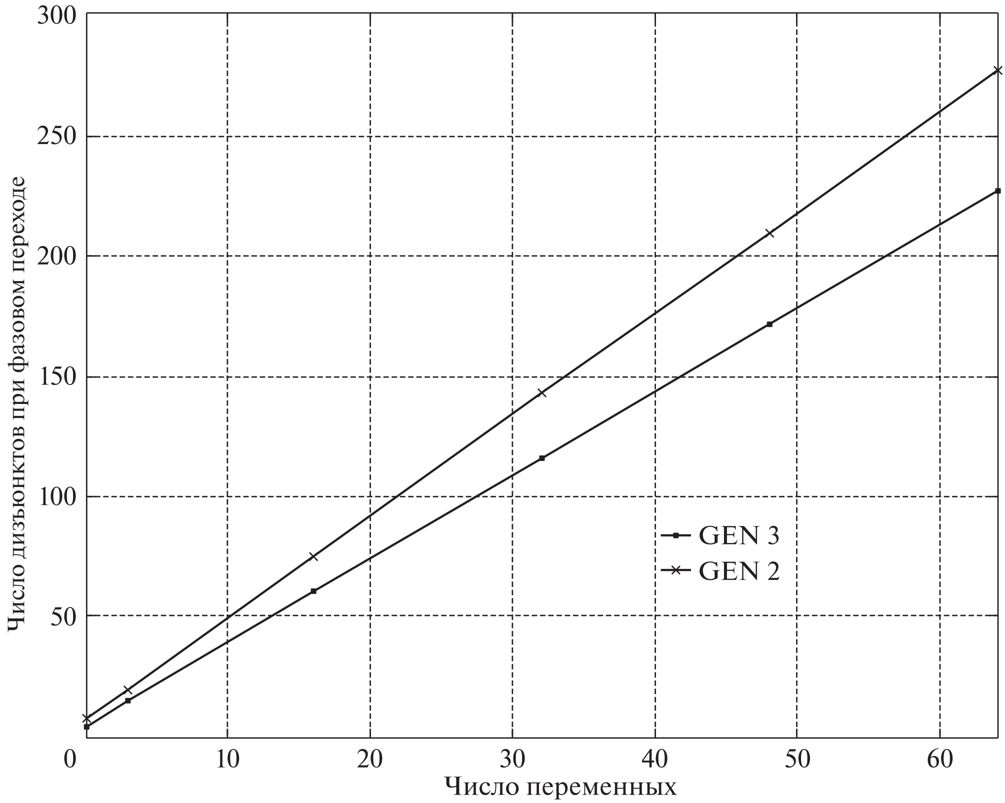

Рис. 4. Иллюстрация эмпирической линейной зависимости $\mathfrak{M}_2$ и $\mathfrak{M}_3$ от N.

асимптотами (при $N \to \infty$) для отношения числа дизъюнктов к числу переменных в точке «фазового перехода» для БПД- и ВНЛ-генераторов соответственно.

В табл. 2 представлены экспериментальные данные для зависимостей $\mathfrak{M}_2(N) = \alpha_2(N)N + 7{,}18$ и $\mathfrak{M}_3(N) = \alpha_3(N)N + 4{,}46$ при числах переменных, превышающих 64. Приведенные результаты позволяют оценить стабильность коэффициентов пропорциональности $\alpha_2(N)$ и $\alpha_3(N)$ при увеличении числа булевых переменных. Каждое из представленных значений получено на основе анализа 200 формул.

Значение коэффициента пропорциональности с вероятностью 0,99 находится в доверительном интервале $(\alpha_i(N) - \delta_i(N), \alpha_i(N) + \Delta_i(N))$, $i = 2, 3$.

Таблица 2. Зависимость коэффициентов пропорциональности от N

N	80	96	112	128	160	192	224	256	288	320
$\alpha_2(N)$	4,23	4,20	4,24	4,22	4,23	4,23	4,21	4,23	4,23	4,23
$\delta_2(N)$	0,04	0,04	0,05	0,04	0,03	0,04	0,02	0,02	0,02	0,02
$\Delta_2(N)$	0,09	0,05	0,03	0,06	0,05	0,04	0,03	0,03	0,03	0,03
$\alpha_3(N)$	3,52	3,49	3,51	3,50	3,51	3,50	3,50	3,49		
$\delta_3(N)$	0,03	0,03	0,02	0,01	0,01	0,02	0,01	0,01		
$\Delta_3(N)$	0,04	0,02	0,01	0,01	0,02	0,01	0,02	0,01		

Невыполнимые формулы полученные ВНЛ-генератором, в среднем существенно сложнее для анализа, чем невыполнимые формулы, полученные БПД-генератором. Анализ формул от 256 переменных, построенных ВНЛ-генератором, длится в среднем в 20 раз дольше, чем для формул, построенных БПД-генератором.

4. Заключение

Предложен и исследован ВНЛ-генератор (выравнивающий нагрузку на литералы) случайных формул 3-КНФ. Этот генератор продуцирует формулы, являющиеся невыполнимыми, при меньшем числе дизъюнктов, чем у широко используемого в настоящее время БПД-генератора (формул без простых дизъюнктов), для которого асимптотическое значение R в точке «фазового перехода» оценивается как 4,23.

Для предложенного в настоящей работе ВНЛ-генератора асимптотическое значение R (при $N \to \infty$) в точке «фазового перехода» оценивается как 3,49.

Характерная для НВЛ,- БПД- и ВНЛ-генераторов линейная зависимость числа дизъюнктов от числа переменных в точке «фазового перехода» удобна при построении тестовых примеров для испытания алгоритмов решения задачи выполнимости (SAT-problem). По заданному числу переменных легко вычисляется количество дизъюнктов, при котором больше половины из генерируемых случайных 3-КНФ формул будут выполнимыми, а при добавлении по одному дизъюнкту в каждую формулу больше половины формул станут невыполнимыми.

Проведенные исследования свидетельствуют в пользу выдвинутой в литературе [5–8] гипотезы о том, что анализ невыполнимых формул, имеющих меньшее число дизъюнктов, как правило, сложнее, чем анализ формул, содержащих больше дизъюнктов. Представляет интерес разработка способов построения формул, невыполнимых при еще меньшем числе дизъюнктов.

Представляет также интерес поиск зависимости количества дизъюнктов, минимально необходимого для невыполнимости нетривиальной 3-КНФ формулы, от числа булевых переменных.

СПИСОК ЛИТЕРАТУРЫ

1. *Biere A., Heule M., Maaren H., Walsh T.* Handbook of Satisfiability // IOS Press, 2009. P. 1–966.
2. *Cook S., Mitchell D.* Finding hard instances of the satisfiability problem: a survey // DIMACS Ser. Discret Math. Theoret. Comput. Sci., Amer. Math. Sos. 1997. V. 35. P. 1–17.
3. *Beame P., Karp R., Pitassi T., Saks M.* On the Complexity of Unsatisfiability Proofs for Random k-CNF Formulas // 30thSTOC, Dallas, TX. May 1998. P. 561–571.
4. *Goldberg A.* On the complexity of the satisfiability problem // Appl. Math. Comput. J. Amer. Math. Sos. 1997. V. 35. No. 1. P. 1–17.
5. *Mitchell D., Selman B., Levesque H.* Hard and easy distribution of SAT problem // Proc. Tenth National Conf. Artific. Intelligence (AAAI-92), San Jose, CA. 1997. P. 459–465.

6. *Crawford J., Auton I.* Experimental Results on the Crossover Point in Satisfiability Problems // Proc. AAAI-93, Washington, DC. 1993. P. 21–27.
7. *Gomes C., Kautz H., Sabharwal A., Selman B.* Satisfiability Solvers / Handbook Knowlege Represent., Elsevier B.V. 2008. P. 88–134.
8. *Heule M.* Minimal unsatisfiable cores of random formulas // Proc. SAT competition. 2013. P. 105.
9. *MohammadTghi Hajighayi, Gregory Sorkin.* The Satisfiability Threshold of Random 3-SAT Is at Least 3.52 // 2003/10/13. arXiv preprint math /0310193.
10. *Kaporis Alexis C., Kirousis Lefteris M., Laias Efthimios G.* The Probabilistic Analysis of a Greedy Satisfiability Algorithm // Random Struct. & Algorithms. 2006. V-28(40). P. 444–480.

Статья представлена к публикации членом редколлегии А.А. Лазаревым.

Поступила в редакцию 06.10.2017

После доработки 15.04.2019

Принята к публикации 18.07.2019