УДК 519.725+004.023

О ВЫЧИСЛЕНИИ МИНИМАЛЬНОГО РАССТОЯНИЯ ДВОИЧНОГО ЛИНЕЙНОГО БЛОЧНОГО КОДА С ИСПОЛЬЗОВАНИЕМ ГЕНЕТИЧЕСКИХ АЛГОРИТМОВ

Могилевская Н.С.

ФГБОУ ВПО «Донской государственный технический университет», Ростов-на-Дону, Россия (344011, Ростов-на-Дону, пл. Гагарина, e-mail: broshka@nm.ru

В работе исследована возможность использования генетических алгоритмов для решения задачи вычисления минимального кодового расстояния линейных блочных кодов. Предполагается, что исследуемые коды не случайные, а получены с использованием различных методов модификации из известных кодов, обладающих хорошими корректирующими свойствами. В работе рассмотрены два известных генетических алгоритма поиска минимального кодового расстояния, а также построен новый алгоритм, который в экспериментальном исследовании показал наилучшие результаты. Сделаны выводы о чувствительности результатов работы исследованных генетических алгоритмов к их настройкам. Показано, что при поиске значения минимального кодового расстояния в случае кодов, заданных не случайным образом, использование нижней алгебраической оценки минимального кодового расстояния предпочтительнее, чем использование для этой цели генетических алгоритмов.

Ключевые слова: минимальное кодовое расстояние, генетический алгоритм, базовые оценки кодов, линейные блочные коды.

CALCULATION OF THE MINIMUM DISTANCE OF BINARY LINEAR BLOCK CODE USING GENETIC ALGORITHMS

Mogilevskaya N.S.

"Don State Technical University", Rostov-on-Don, Russia (344000, Rostov-on-Don, pl. Gagarin, e-mail: broshka@nm.ru

We have studied the possibility of using genetic algorithm for solving the problem of computing the minimum code distance of linear block codes. It was assumed that the analyzed codes are not random, but obtained using different methods of modification of the known codes having good correcting properties. The paper considers two known genetic algorithm of search the minimum code distance, and also built a new algorithm, which in the experimental study showed the best results. Conclusions are made about the sensitivity of the results is investigated genetic algorithms from their settings. It is shown that when searching for a minimum code distance in the case of codes defined not random, use the lower algebraic estimates of the minimum code distance is preferable to use for this purpose genetic algorithms.

Keywords: minimum distance, genetic algorithm, the basic assessment codes, linear block code.

Линейным блочным $(n,k)_q$-кодом C называют подмножество линейно-векторного пространства F_q^n размерности k (k<n), где F_q – поле Галуа мощностью q [3, 5]. Одним из важных параметров линейных кодов является минимальное кодовое расстояние $d_{\min}$ (МКР), определяемое расстоянием между кодовыми словами в метрике Хемминга, и определяющее способность кода по исправлению и обнаружению ошибок. Значение МКР произвольного блочного $(n,k)_q$-кода C можно вычислить по формуле [3, 5]:

$$d_{\min} = \min\{wt_h(\bar{c}_i) \,|\, \bar{c}_i \in C, \bar{c}_i \neq \bar{0}\},$$

где $\bar{c}_i \in F_q^n$ - кодовое слово кода C, функция $wt_h(\bar{c}_i)$ возвращает вес Хемминга слова $\bar{c}_i$, т.е. число ненулевых позиций в этом векторе. Один из способов задания кода состоит в использовании порождающей матрицы G кода C, в этом случае оператор кодирования

$$C : F_q^k \to F_q^n \tag{1}$$

определен формулой $\bar{i} = \bar{c}G$, где $\bar{i}$ – информационное слово, а $\bar{c}$ – кодовое слово.

Нахождение минимального кодового расстояния для произвольных кодов является NP-полной задачей [8], а поиск эффективных алгоритмов вычисления $d_{\min}$ – открытая проблема в теории кодирования. МКР является важным параметром, не только для приложений, использующих помехоустойчивое кодирование для обеспечения достоверности передаваемых данных, но и в ряде криптографических приложений. Отметим, что чем больше значение $d_{\min}$, тем лучшими корректирующими способностями обладает помехоустойчивый код. Для небольшого числа кодов, в основе которых лежит некоторая комбинаторная или алгебраическая структура найдены аналитические формулы для вычисления МКР, примерами таких кодов являются коды Хемминга, Рида-Маллера, Рида-Соломона. У большинства кодов в порождающей матрице отсутствует какая-либо структура, такие коды называют кодами общего положения [5]. Задать такой код можно, например, с помощью порождающей матрицы, выбранной случайным образом среди всех матриц определенных размерностей.

В работе [6] предложено использовать эвристические алгоритмы для поиска значения МКР линейного кода, заданного порождающей матрицей G.

Цель настоящей работы состоит в проведении исследования по оценке эффективности использования генетических алгоритмов (ГА) для нахождения МКР не случайного двоичного линейного помехоустойчивого кода и получении вывода о целесообразности использования генетических алгоритмов для решения задачи поиска $d_{\min}$.

Общая идея использования генетических алгоритмов в задаче поиска МКР

Генетические алгоритмы относятся к стохастическим, эвристическим методам поиска решений задач оптимизации и моделирования путём случайного подбора, комбинирования и вариации искомых параметров с использованием механизмов, аналогичных естественному отбору в природе, см., например [4]. Уточним основные понятия, используемые в теории ГА, применительно к решаемой задаче поиска МКР линейного блочного $(n,k)_2$-кода C. В качестве особи, входящей в популяцию, будем рассматривать информационные слова кода C, которые представляют собой векторы $\bar{a} = (a_1,...,a_k) \in F_2^k$, заданные над полем F_2. Геном назовем символ $a_i \in F_2$ особи $\bar{a}$. Пусть $f = wt_h(C(\bar{a}))(\in N \cup \{0\})$, где wt_h функция веса Хемминга, C – оператор кодирования (1), тогда функцию приспособленности (фитнес-функцию) определим следующим образом:

$$fitness(\bar{a}) = \begin{cases} f, если\ f \neq 0 \\ n, если\ f = 0 \end{cases}. \tag{2}$$

Оператор мутации $\mathrm{mutate}(s, p_m, z)$ с вероятностью p_m инвертирует значения z различных генов особи s. Оператор скрещивания $\mathrm{cross}(s_1, s_2, p_c, z)$ с вероятностью p_c репродуцирует особей-родителей s_1 и s_2 и порождает двух особей-потомков ch_1 и ch_2. Если согласно вероятности скрещивание должно произойти, то случайным образом определяется z различных точек скрещивания l_z, где $l_z \in [1..k-1]$. В данной работе рассматривается одно- и двухточечное скрещивание. При одноточечном скрещивании, у одного из потомков на позициях от 1 до l_z которого стоят гены первого родителя, а на позициях от l_z+1 до k стоят гены второго родителя, у второго потомка, на позициях от 1 до l_z которого стоят гены второго родителя, а на позициях от l_z+1 до k стоят гены первого родителя. В случае двухточечного скрещивания потомки наследуют фрагменты наборов родительских генов, определяемые двумя случайно выбранными точками скрещивания.

Генетические алгоритмы нахождения минимального кодового расстояния

Ниже в работе рассмотрены два генетических алгоритма, обозначенные в данной работе как алгоритм А и алгоритм Б, предложенные в работе [6], а также новый генетический алгоритм поиска минимального кодового расстояния линейного блочного кода, обозначенный далее, как алгоритм В.

На вход всех алгоритмов поступают параметры $(n,k)_q$-кода C, определенного порождающей матрицей G, параметры операторов мутации и скрещивания, M – число особей в популяции, N_{max} – число популяций, которое необходимо сгенерировать. На выходе алгоритмов формируется значение, которое предположительно является минимальным кодовым расстоянием кода C. В алгоритме Б в список входных параметров дополнительно подается число элитных особей N_e. Результатом работы алгоритмов является наименьшее значение фитнес-функции особей из поколения с номером N_{max}.

Рассмотрим алгоритм А [6]. Начальная популяция формируется случайным образом из M особей вида $\bar{a} \in F_2^k$, для каждой особи вычисляется функция приспособленности (2). Затем генерируется N_{max} популяций, при этом каждая популяция N_i строится из предыдущей популяции $N_{(i-1)}$ по следующей схеме. Особи поколения $N_{(i-1)}$ сортируются в порядке возрастания значений их функций приспособленности. В поколение N_i включаются M/2 особей с наименьшим значением функции (2) из популяции $N_{(i-1)}$. Для генерации оставшихся M/2 особей из особей поколения $N_{(i-1)}$ случайным образом выбираются две особи s_1 и s_2, к каждой из которых применяется оператор мутации, а затем одноточечный оператор скрещивания:

$$s_1' = \mathrm{mutate}(s_1, p_m, n),\ s_2' = \mathrm{mutate}(s_2, p_m, n),\ (ch_1, ch_2) = cross(s_1', s_2', p_c, z).$$

Потомок с наименьшим значением функции приспособленности, включается в поколение N_i.

В алгоритме Б [6] начальная популяция генерируется случайным образом из M особей вида $\bar{a} \in F_2^k$, для каждой особи вычисляется функция (2). Следующие N_{max} популяций генерируются рекурсивно. Особи популяции $N_{(i-1)}$ сортируются в порядке возрастания значений их функций приспособленности. В поколение N_i включаются N_e особей с наименьшим значением фитнес-функции из популяции $N_{(i-1)}$. Для генерации оставшихся (M-N_e) особей с помощью турнирного отбора из поколения $N_{(i-1)}$ выбираются две особи s_1 и s_2 для репродукции, с вероятностью p_c эти особи скрещиваются $(ch_1, ch_2) = \text{cross}(s_1, s_2, p_c, 2)$, к потомкам применяется функция мутации $ch_1' = \text{mutate}(ch_1, p_m, k)$, $ch_2' = \text{mutate}(ch_2, p_m, k)$, затем оба потомка переходят в следующее поколение.

Первая особь начальной популяции алгоритма В формируется в виде нулевого вектора из F_2^k, остальные (M-1) особей генерируются случайно в виде $\bar{a} \in F_2^k$, для каждой особи, кроме первой, вычисляется функция (2). Для построения следующего поколения особи популяции $N_{(i-1)}$ сортируются в порядке возрастания значений их фитнес-функций. В поколение N_i включаются нулевая особь и две особи с наименьшим значением функции (2) из $N_{(i-1)}$. Для получения остальных особей нового поколения выполняются следующие действия. Для каждой новой s_i (i=2,3,4,…) особи из формируемого поколения N_i выбирается случайным образом особь s_j для скрещивания $(ch_1, ch_2) = \text{cross}(s_i, s_j, p_c, 1)$. Потомок с максимальным значением фитнес-функции отбрасывается, к потомку ch с минимальным значением функции приспособленности применяется оператор мутации: $ch' = \text{mutate}(ch, p_m, 1)$. Из особей поколения N_i случайным образом выбирается еще одна особь a. Для особей a и ch' вычисляется значение функций приспособленности, затем особь, с наименьшим значением функции (2) включается в формируемое поколение.

Экспериментальное исследование

Описанные алгоритмы реализованы с использованием математического пакета Matlab. В работе исследовано 38 различных кодов, длина n которых варьируется от 5 до 512. В этот набор вошли как случайные коды, так и известные коды Хемминга, Голея, БЧХ, а также модификации известных кодов методами укорочения, перфорации, расширения и др. Описание использованных методов можно найти в работах [1], [3], для выполнения различных модификаций рассматриваемых кодов использовано специализированное программное средство [2]. Реальное значение МКР получено с использованием переборного алгоритма, в ходе которого вычислялись веса всех кодовых слов.

В одной из серий экспериментов исследованы БЧХ-коды, использованные в [6] с применением указанных там же параметров алгоритмов А и Б. Среднее значение разницы между результатами, полученными в данной работе, и данными из [6] составляет 29% от

значения истинного минимального кодового расстояния. Полагаем, что такое расхождение не противоречит эвристической природе использованных алгоритмов.

В табл. 1 представлены результаты поиска минимального кодового расстояния рассмотренными выше генетическими алгоритмами для некоторых кодов, случайным образом выбранных из набора исследуемых кодов. Отметим, что значения, указанные в таблице, являются лучшими результатами, полученными при различных размерах начальной популяции, а так же различных вероятностях мутации и кроссовера. Структура таблицы следующая: в первом столбце указан исследуемый код и его параметры в виде тройки (n,k,d_{min}), второй-четвертый столбцы содержат значения минимального кодового расстояния, полученные с использованием генетических алгоритмов А, Б из [6] и нового алгоритма В, соответственно. Содержимое последнего столбца будет рассмотрено позже.

Таблица 1

Значение МКР, найденного генетическими алгоритмами А, Б и В

Код	Минимальное кодовое расстояние			
	Алгоритм А	Алгоритм Б	Алгоритм В	Оценки
(7,4,3)	3	3	3	3..3
(15,6,6)	6	6	6	6..6
(23,12,7)	7	7	7	7..7
(63,51,5)	12	14	5	5..5
(63,51,2)	2	2	2	5..5
(127,64,21)	24	27	21	21..28
(127,92,10)	13	16	12	11..14
(127,113,5)	24	28	7	5..5
(255,71,59)	73	90	67	61..89
(255,71,8)	8	9	8	61..89
(255,223,9)	12	16	11	9..10

Из табл. 1 видно, что алгоритм Б показал наихудшие, а новый алгоритм В наилучшие результаты, аналогичные результаты справедливы и для других исследованных кодов.

Результаты работы генетических алгоритмов поиска минимального кодового расстояния значительно зависят от параметров алгоритмов, а именно, от размера и числа популяции, вероятностей кроссовера и мутации. Так полученные данные демонстрируют понижение точности при повышении вероятности мутации, что может объясняться повышающейся вероятностью спонтанного наделения отбираемых особей плохими свойствами. Так же можно говорить о том, что точность результата определяется размерами

пространства поиска: для повышения точности следует увеличивать размер исходной популяции, что, в свою очередь повышает время выполнения алгоритма. Выявить четкую зависимость эффективности рассматриваемых алгоритмов от диапазона значений того или иного стохастического оператора не удалось в силу нестабильности полученных результатов.

Точный результат исследуемые ГА выдают всегда в случае, когда рассматриваемый код содержит небольшое число кодовых слов. Этот факт имеет простое объяснение, пусть, например, в ГА формируется 400 популяций из 1000 особей. При поиске МКР для кода (7,4,3), содержащего всего 2^4=16 кодовых слов, с большой долей вероятности в начальной популяции слово минимального веса будет сформировано. Более того, для этого кода поиск МКР прямым перебором окажется быстрее. Однако, если применить эти же настройки параметры для (63,51,5)-кода, содержащего $2^{51}=(2^3)^{17}\approx10^{17}$кодовых слов, то при поиске МКР вероятность генерации в начальной популяции слова минимального веса резко падает.

Проведенное исследование выявило следующие недостатки изучаемых генетических алгоритмов: длительное время работы, возможную вырождаемость решений, невысокую точность.

На вход исследуемых генетических алгоритмов поступали коды, заданные порождающими матрицами, генерация которых производилась как случайным образом (коды (63,51,2) и (255,71,8)), так и не случайным образом (все остальные коды из табл. 1). В случае случайных кодов результаты работы ГА достаточно точные. Однако в практических приложениях коды, имеющие большую избыточность и малую корректирующую способность, не используются. При использовании помехоустойчивых кодов в реальных приложениях, как для защиты данных от помех, так и в криптографических задачах, используют коды, обладающие хорошими корректирующими способностями. При использовании таких кодов следует ожидать, что значение минимального кодового расстояния будет достаточно большим для заданных длины и размерности кода.

Алгебраическая оценка минимального кодового расстояния

В теории помехоустойчивого кодирования известен ряд оценок, связывающих параметры линейных блочных кодов. К таким оценкам относятся, например, хорошо известные границы Хемминга, Варшамова-Гилберта, Синглтона, Бассалыго-Элайеса, Грайсмера и другие [3], [5]. Используя данные оценки для $(n,k)_q$-кода можно оценить верхнюю и нижнюю границы значения минимального кодового расстояния $(n,k)_q$-кода. Рассмотрим, например, границы Хемминга и Гилберта. По отношению друг к другу граница Хемминга является «верхней», а граница Гилберта «нижней». Так, если произвольно выбранные параметры кода n, k и d_{min} не удовлетворяют границе Хемминга, то кода с такими параметрами не существует. Если параметры кода удовлетворяют границе Гилберта, то код

с такими параметрами существует. Если же выбранные параметры удовлетворяют границе Хемминга, но не удовлетворяют границе Гилберта, то вопрос о существовании такого кода не решен полностью (не смотря на множество частных результатов) [5].

Электронный ресурс [7], посвященный линейным блочным кодам, позволяет для введенных значений длины n, размерности k и мощности q кода вычислить нижнюю и верхнюю оценки МКР кода с использованием целого ряда известных оценок. Для всех кодов, использованных в исследовании генетических алгоритмов, были вычислены такие оценки минимального кодового расстояния кода (см. последний столбец табл. 1). Для кода длиной 127 и размерностью 64 нижняя оценка равна 21, а верхняя – 28, следовательно, использованный в экспериментах (127,64,21)-код можно назвать «хорошим», т.к. его МКР лежит на нижней границе, т.е. значение $d_{min}=21$ является максимальным для которого, согласно базовым оценкам линейных блочных кодов, гарантировано существует код с указанными параметрами длины и размерности. Для $(255,71)_2$-кода базовые оценки гарантируют существование кода с $d_{min}=61$, а у кода, использованного в эксперименте это значение меньше и равно 59, это говорит о том, что можно построить $(255,71)_2$-код с большей корректирующей способностью.

Анализ реального значения минимального кодового расстояния исследуемых кодов, значений МКР, найденных с использованием генетических алгоритмов, а также значений нижней оценки МКР, позволяет сделать вывод о том, что для кодов, чьи порождающие матрицы заданы не случайно, а обладают некоторой полезной комбинаторной или алгебраической структурой, вместо использования генетических алгоритмов для поиска минимального кодового расстояния целесообразно использовать нижнюю оценку МКР. В этом случае погрешность между реальным значением МКР и его нижней оценкой не превосходит погрешность между реальным значением МКР и значением, найденным с использованием генетического алгоритма. К тому же время вычисления оценки значительно меньше времени работы генетического алгоритма.

Заключение

В работе построен новый генетический алгоритм поиска минимального кодового расстояния линейного блочного кода, заданного порождающей матрицей. Результаты экспериментов показали, что новый алгоритм работает эффективнее алгоритмов из [6], взятых за основу. Однако все рассмотренные в работе алгоритмы выдают результат с невысокой точностью. В работе показано, что в случае двоичных линейных блочных кодов, порождающие матрицы которых построены не случайно, использование нижней оценки минимального кодового расстояния предпочтительнее применения описанных генетических алгоритмов с точки зрения времени работы и точности результата.

Список литературы

1. Могилевская Н.С., Сухоставская К.С. Об экспериментальном исследовании характеристик модифицированных помехоустойчивых блочных двоичных кодов // Вестник Донского гос. техн. ун-та. – 2007. – Т.7. – №3. – С. 276-282.

2. Могилевская Н.С., Шпыгарь С.М. Свидетельство о государственной регистрации программы для ЭВМ №2009615850 Российская Федерация. Программа модификации блочных линейных кодов «New Code». Зарегистр. 12.07.2009. – 24 с.

3. Морелос-Саргоса Р. Искусство помехоустойчивого кодирования. Методы, алгоритмы, применение. – М.: Техносфера, 2005. – 320 с.

4. Рутковская Д., Пилиньский М., Рутковский Л. Нейронные сети, генетические алгоритмы и нечеткие системы. – М.: Горячая линия – Телеком. – 2006. – С. 124-170.

5. Сидельников В.М. Теория кодирования. – М.: ФИЗМАТЛИТ. – 2008. – 324 с.

6. Askali M., Azouaoui A., Nouh S., Belkasmi M. On the Computing of the Minimum Distance of Linear Block Codes by Heuristic Methods // International Journal of Communications, Network and System Sciences. – 2012. – № 5. – P. 774-784.

7. Grassl M. Bounds on the minimum distance of linear codes and quantum codes // URL: http://www.codetables.de. (accessed on 26.11.2014).

8. Vardy A. The intractability of computing the minimum distance of a code // IEEE Trans. Inf. Theory, vol. 43, no. 6, pp. 1757–1766, 1997.

Рецензенты:

Габриэльян Д.Д., д.т.н., профессор, заместитель начальника научно-технического комплекса «Антенные системы» по науке, Федеральный научно-производственный центр ФГУП «РНИИРС» г. Ростов-на-Дону;

Звездина М.Ю., д.ф.-м.н., доцент, зав. кафедрой «Радиоэлектроника», Минобрнауки России, ФБГОУ ВПО «Донской государственный технический университет», г. Ростов-на-Дону.