

УДК 004.891.021

UDC 004.891.021

**ИДЕНТИФИКАЦИЯ МЕХАНИЗМОВ
РЕАЛИЗАЦИИ ОПЕРАТОРОВ
ГЕНЕТИЧЕСКОГО АЛГОРИТМА В
ЭКСПЕРТНЫХ СИСТЕМАХ
ПРОДУКЦИОННОГО ТИПА**

**IDENTIFICATION OF MECHANISMS OF
REALIZATION OF GENETIC ALGORITHM
OPERATORS IN EXPERT SYSTEMS OF
PRODUCTION TYPE**

Частикова Вера Аркадьевна
к.т.н., доцент
*Кубанский государственный технологический
университет, Краснодар, Россия*

Chastikova Vera Arkadyevna
Cand.Tech.Sci., associate professor
*Kuban State Technological University, Krasnodar,
Russia*

В статье проводится идентификация и исследование ключевых параметров генетического алгоритма метода генетических схем и их влияние на эффективность поиска решений в экспертных системах производственного типа. Рассматриваются следующие параметры генетического алгоритма: оператор кроссовера, выбор родительской пары, оператор мутации, оператор инверсии

In this article the identification and the research of genetic algorithm key parameters of genetic schemes method and their influence on efficiency of optimum decisions search in expert systems of production type is conducted. The following parameters of genetic algorithm are considered: crossover operator, choice of parents pare, mutation operator, inversion operator

Ключевые слова: ГЕНЕТИЧЕСКИЙ АЛГОРИТМ, ЭКСПЕРТНАЯ СИСТЕМА, БАЗА ЗНАНИЙ, ОПЕРАТОР КРОССОВЕРА, ОПЕРАТОР МУТАЦИИ, РОДИТЕЛЬСКАЯ ПАРА, ОПЕРАТОР ИНВЕРСИИ, ГИПОТЕЗА

Keywords: GENETIC ALGORITHM, EXPERT SYSTEM, KNOWLEDGE BASE, CROSSOVER OPERATOR, MUTATION OPERATOR, PARENTS COUPLE, INVERSION OPERATOR, HYPOTHESIS

В разработанном программном исследовательском комплексе (ПИК) «Поиск» [2,3] для изучения эффективности использования теоретически обоснованного и разработанного автором метода генетических схем (ГС) возможна идентификация и исследование влияния различных параметров генетического алгоритма (ГА) на эффективность поиска решений в экспертных системах производственного типа. В настоящее время теория генетических алгоритмов и эволюционного программирования является довольно развитой, [3] и для каждого из операторов ГА предложено различное количество модификаций. Идентификация механизмов реализации операторов генетического алгоритма подразумевает под собой выбор наиболее оптимальных модификаций ключевых параметров ГА для поиска решений в интеллектуальных системах, основанных на знаниях. В данной статье рассматриваются следующие основные параметры генетического алгоритма метода генетических схем:

- оператор кроссовера;
- способ формирования родительской пары;
- оператор мутации;
- оператор инверсии.

В качестве целевой функции (ЦФ) принимается число гипотез, доказанных вхолостую [4,5].

Выбор типа оператора кроссовера. В качестве генетических операторов получения новых генотипов потомков возможно применение двух типов кроссовера:

- однотоочечный кроссовер;
- многотоочечный кроссовер.

В качестве многотоочечного используется побитовый кроссовер, где очередная точка разрыва хромосом определяется по наступлению с заданной вероятностью события.

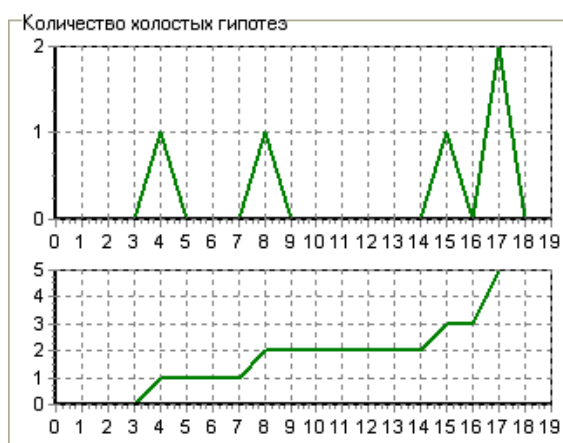
Для исследований возможна установка и такого режима, который использует в репродукции вместо двух только одного потомка.

Вычислительные эксперименты, проведенные автором статьи [1, 2], также показали, что в подавляющем большинстве лучшие результаты даже для простых случаев получены для многотоочечного кроссовера. Преимущества выбора побитового оператора, естественно, сильнее проявляются при работе с базами знаний со сложной структурой ландшафта функции приспособленности.

Представленные на рисунке 1 диаграммы иллюстрируют именно тот случай, когда лучшие результаты получены для побитового кроссовера в работе с тестовыми деревьями, даже не отличающимися особой сложностью. Такой результат получен для дерева базы знаний (БЗ) (15*4) с числом брачных пар равным 100, при 300 особях в популяции на протяжении 20-ти последовательных экспериментов. Причем необходимо

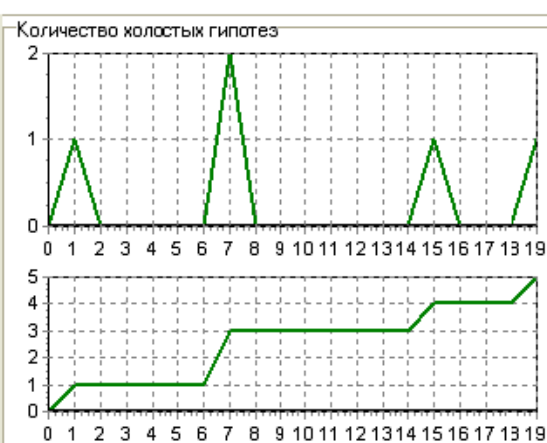
обратить внимание на тот факт, что получены схожие результаты для двух схем размножения.

В то же самое время эксперименты подтвердили тот факт, что при работе с деревьями простой структуры нельзя однозначно отдавать предпочтение одно- или многоточечному кроссоверу. Подчас результаты использования одноточечного кроссовера в работе с некоторыми тестовыми деревьями оказывались более эффективными [3].

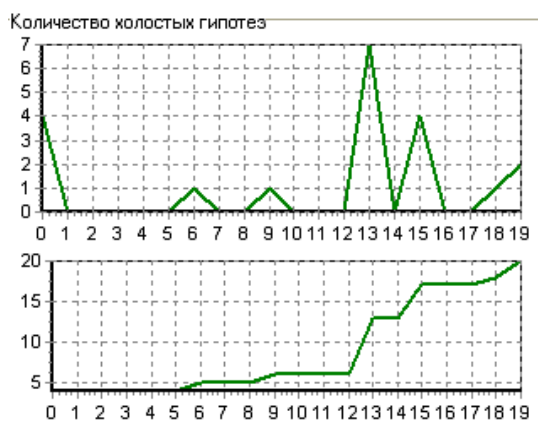


детерминированная схема

а) побитовый кроссовер

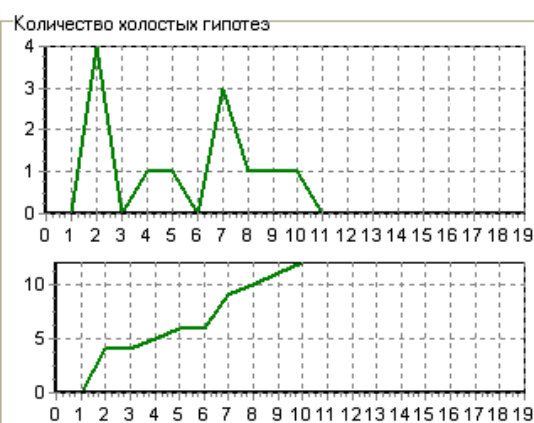


вероятностная схема



детерминированная схема

б) одноточечный кроссовер



вероятностная схема

Рисунок 1 – Сравнительная оценка двух видов кроссовера ГА

Для последнего случая на рисунке 2 приведены изменения лучшей, средней и худшей степеней достоверности для 3-ей гипотезы на протяжении всех итераций работы ГА.

Эксперименты показали, что использование механизма случайного выбора типа кроссовера для каждой конкретной брачной пары очень часто оказывается более эффективным, чем детерминированный подход к выбору типа кроссовера.

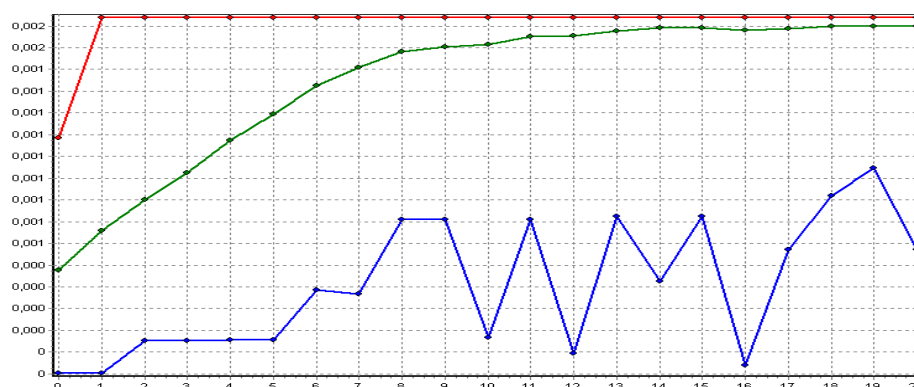


Рисунок 2 – Лучшая, средняя и худшая степени достоверности для 3-ей гипотезы на протяжении работы ГА

Случайный выбор позволяет сгладить различия этих двух подходов и улучшить показатели среднего ожидаемого результата. Это происходит из-за того, что достаточно трудно для деревьев БЗ с неизвестным поведением целевых функций априорно определить – который из двух операторов более подходит для каждого конкретного ландшафта приспособленности. Для всех проведенных тестов так и произошло: применение случайного механизма в выборе кроссовера дало лучшие результаты по сравнению с детерминированными подходами.

ПИК «ПОИСК» предлагает различные возможности для углубленного исследования детерминированного метода, а именно того, что касается выбора родительской пары [1,3].

Выбор способа формирования родительской пары. При фиксированном числе брачных пар существует несколько подходов выбора возможных пар для репродукции:

- случайный выбор родительской пары;
- селективный подбор родителей;
- инбридинг;
- аутбридинг.

В ПИК «ПОИСК» первые два подхода не рассматриваются по ряду причин, главные из которых определяются следующими положениями.

Случайный выбор родительской пары – прост, универсален для решения различных классов задач. Однако он достаточно критичен к численности популяции, поскольку эффективность алгоритма, реализующего такой подход, снижается с ростом численности популяции.

Селективный отбор, предполагающий, что родителями могут стать только те особи, значение приспособленности которых не меньше среднего значения приспособленности по популяции, обеспечивает более быструю сходимость алгоритма. Однако из-за быстрой сходимости селективный выбор родительской пары не подходит для многоэкстремальных задач со сложным ландшафтом приспособленности. Этот недостаток должен быть компенсирован использованием подходящего механизма отбора, который бы тормозил слишком быструю сходимость алгоритма.

В качестве механизмов регулирования отбора родительских пар в ПИК «ПОИСК» используются:

- инбридинг;
- аутбридинг.

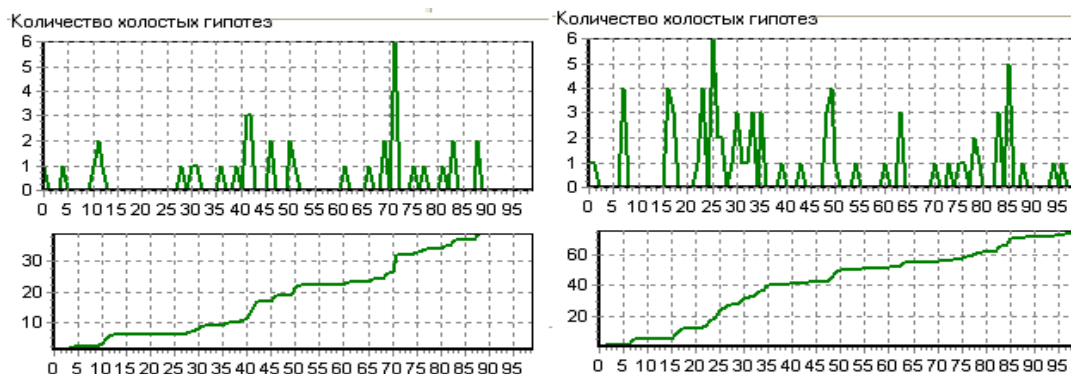
Инбридинг построен на формировании пары на основе близкого «родства». Под инбридингом понимается такой метод, когда первый член пары выбирается случайно, а вторым с большей вероятностью будет

максимально близкая к нему по приспособленности особь. В ПИК «ПОИСК» для реализации такого механизма отбора родительских пар на панели помещено специальное поле, с помощью которого задается интервал поиска родителей в списке хромосом. Например, если данное поле содержит число 20, то это означает, что для скрещивания в качестве первого родителя берется первая хромосома, а в качестве второго выбирается самая близкая по приспособленности хромосома из следующих за ней 20-ти, для второй хромосомы поиск самой близкой хромосомы ведется среди следующих за ней 20-ти и т. д.

Аутбридинг построен на формировании пары на основе дальнего «родства» и формирует брачные пары из максимально далеких особей [1]. Именно аутбридинг призван препятствовать преждевременной сходимости к некоторому квазиоптимальному решению в процессе поиска. Естественно, что применение этого механизма увеличит время поиска [3].

Экспериментальные исследования показали, что неплохие результаты дает сочетание обоих механизмов. Такое сочетание возможно и в ПИК «ПОИСК». Механизм аутбридинга требует задания в расположенном рядом с его флажком поле периода его выполнения, а в поле, расположенном ниже под названием «Количество хромосом для проверки аутбридинга», введения значения интервала, на котором будут отыскиваться подходящие хромосомы. Пример: в эти поля помещены числа 8 и 50. Это означает, что аутбридинг будет выполняться над выделенной частью списка хромосом, длиной в 50 хромосом, с периодом 8: 7-мь поколений выполняется в соответствии со своими установками инбридинг, в 8-ом поколении выполняется аутбридинг.

Ниже для параметров кроссовера, озвученных выше, приведены результаты экспериментов при чистом инбридинге и совместном использовании инбридинга и аутбридинга (рис. 3).



а) чистый инбридинг

б) совместное использование инбридинга и аутбридинга

Рисунок 3 – Результаты эксперимента для различных механизмов регулирования отбора родительских пар ГА

Наиболее полезно совместное применение обоих представленных методов для многоэкстремальных задач. Однако два этих способа по-разному влияют на поведение генетического алгоритма. Так инбридинг можно охарактеризовать свойством концентрации поиска в локальных узлах, что фактически приводит к разбиению популяции на отдельные локальные группы вокруг подозрительных на экстремум участков ландшафта, напротив, аутбридинг как раз направлен на предупреждение сходимости алгоритма к уже найденным решениям, заставляя алгоритм просматривать новые, неисследованные области.

Влияние на эффективность поиска типа используемого оператора мутации. Параметры оператора мутации устанавливаются с помощью показанной ниже на рисунке 4 панели ПИК «ПОИСК».

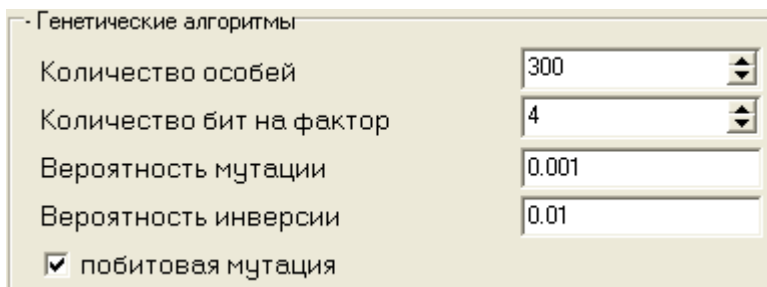


Рисунок 4 – Панель установки значений параметров оператора мутации ГА

В ПИК «ПОИСК» реализована однобитовая мутация. Устанавливаемая с помощью этой панели вероятность мутации относится к каждому биту каждой хромосомы. В силу традиционно малой вероятности мутации (в экспериментах вероятность мутации составляла 0.001 - 0.1) влияние ее на процесс поиска невелико [1].

Ниже на рисунке 5 помещена диаграмма влияния величины мутации на эффективность проведения консультаций. Наилучшие результаты, согласно данной диаграмме, были получены при вероятности мутации равной 0.01. Для этого значения вероятности мутации на рисунке 6 приводятся две диаграммы, характеризующие сам процесс консультаций.

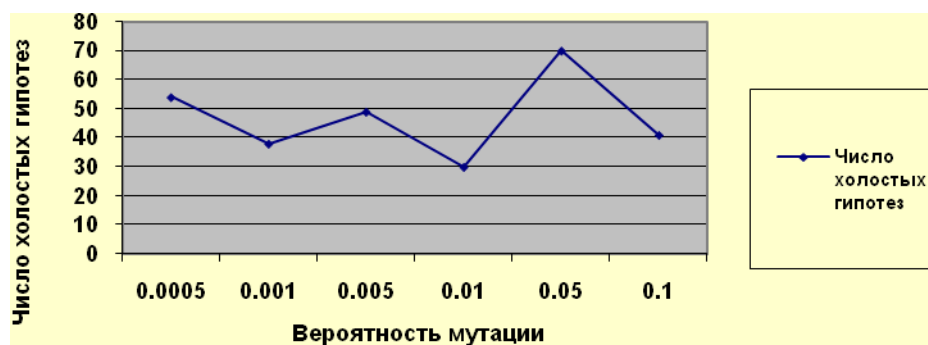


Рисунок 5 – Результаты эксперимента с изменяемыми значениями вероятности мутации

Как отображено на верхней диаграмме, данный процесс характеризуется практически плотным попаданием в процессе поиска оптимальных решений на сгенерированные схемы третьего метауровня базы знаний, что в конечном итоге означает плотное покрытие схемами удовлетворительных точек ландшафта функции приспособленности.

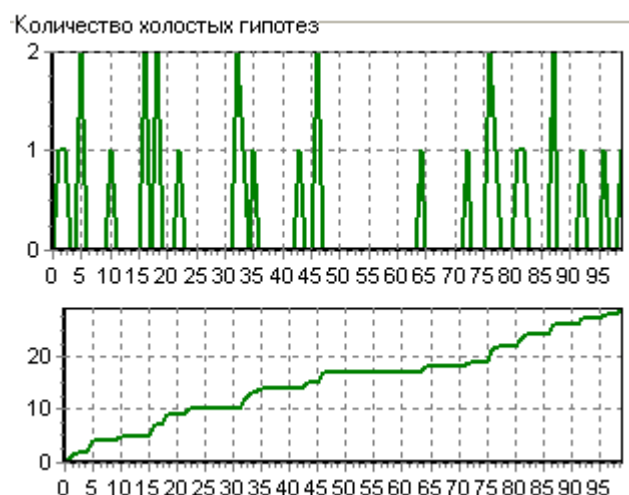


Рисунок 6 – Результаты эксперимента со значением вероятности мутации равным 0.01

Операторы мутации призваны порождать новые особи с отличными от уже имеющихся в популяции особей свойствами. Они, также как и механизм аутбридинга, предотвращают раннюю сходимость к некоторому экстремуму процесса поиска, а поэтому их роль неопределима в случае сложного ландшафта целевой функции. Зависимость величины степени достоверности гипотезы от степеней достоверностей терминальных фактов имеет многокритериальный характер, и неизвестный ландшафт ее поведения предполагает активное использование операторов мутации.

Влияние оператора инверсии на эффективность поиска. На вышеприведенной панели (рис. 4) существует еще одно поле для задания вероятности выполнения оператора инверсии. Операторы инверсии, также как и операторы мутации, призваны порождать новые особи с отличными от уже имеющихся в популяции свойствами и предназначены предотвращать преждевременную сходимость процесса поиска, побуждая ГА включать в новое поколение образовавшиеся новые индивидуумы, вновь расширяя тем самым область поиска [2,3].

На рисунке 7 для дерева БЗ с 15-ю гипотезами и 4-мя уровнями представлена диаграмма, иллюстрирующая поведение метода ГС в тех

случаях, когда метауровни базы знаний формируются при различных значениях вероятностей выполнения оператора инверсии.

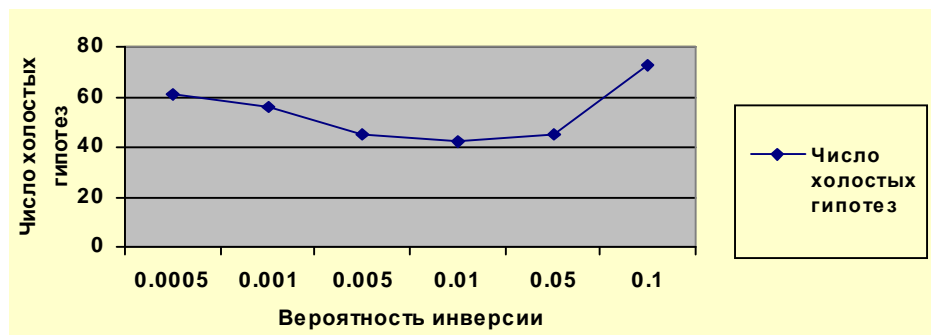


Рисунок 7 – Результаты эксперимента при вариациях вероятности инверсии

На рисунке 8 приведены результаты консультаций для сформированной базы метазнаний в течение 100 экспериментов при вероятности инверсии, равной 0.01, для которой диаграмма 7 дает наилучшие результаты. Как видно на верхней диаграмме рисунка 8, в 91 случае из 100 процесс поиска в эксперименте завершается либо точным попаданием на достоверную гипотезу, либо после доказательства вхолостую только одной из них.

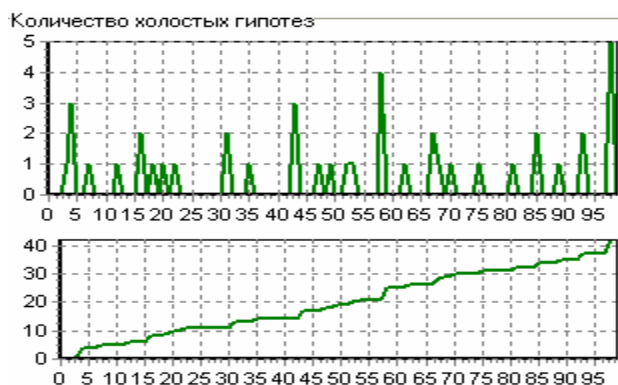


Рисунок 8 – Результаты эксперимента со значением вероятности инверсии равным 0.005

Поведение метода ГС для различных видов деревьев БЗ (дерево 1 - 15*4, дерево 2 - 20*4, дерево 3 - 20*5) при изменении вероятности выполнения оператора инверсии в пределах от 0.001 до 0.1 поясняется результатами, помещенными в таблице 1 и на рисунке 9.

Таблица 1 – Влияние величины вероятности инверсии на количество холостых гипотез

Число холостых гипотез		Вероятность инверсии				
		0.1	0.05	0.01	0.005	0.001
	Дерево 1	24	21	5	6	12
	Дерево 2	20	14	20	17	40
	Дерево 3	8	14	17	8	5
	Среднее значение	17,33	16,33	14	10,33	19

Многочисленные результаты экспериментов, в том числе представленные и на диаграмме рисунка 9, не позволяют сделать однозначные выводы о влиянии величины вероятности инверсии на эффективность поиска [1]. Поведение кривых графиков говорит о том, что это влияние всецело зависит от структуры дерева БЗ, а значит опять-таки от ландшафта функции приспособленности, и при неизвестном поведении последней подходящая величина вероятности инверсии должна подбираться экспериментально. Тем не менее, стоит отметить для приведенных зависимостей значение вероятности инверсии равное 0.005, при котором усредненное по всем графикам число гипотез, доказанных вхолостую, дает наименьшее значение.

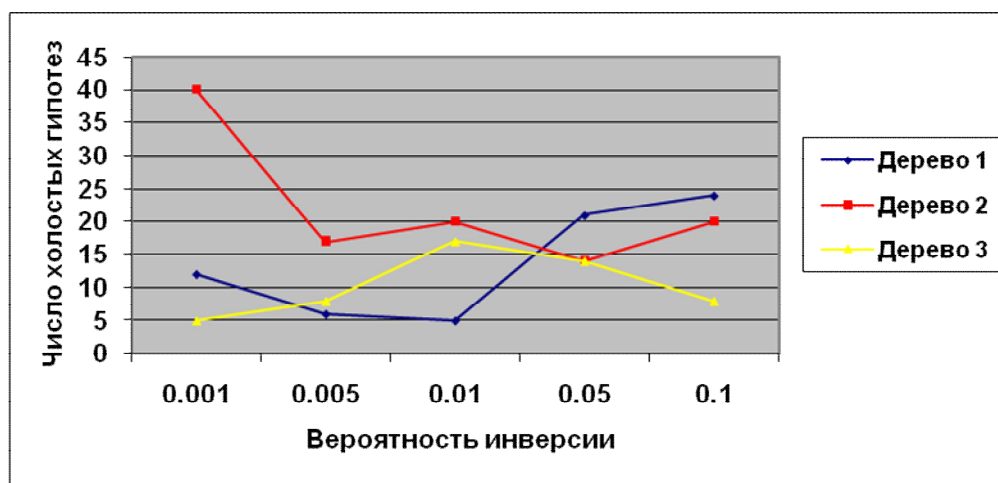


Рисунок 9 – Диаграмма влияния величины вероятности инверсии на эффективность поиска

Рассмотрев и проанализировав в отдельности пути предотвращения преждевременной сходимости процесса поиска, реализованные в ПИК для метода ГС, такие как:

- кроссовер с аутбридингом,
- оператор мутации,
- оператор инверсии,

следует оценить степень их совместного влияния на эффективность работы метода ГС.

В таблице 2 и на рисунке 10 приведены кривые влияния вероятности инверсии на эффективность процесса поиска при различных сочетаниях указанных выше факторов. Причем приведены результаты для такого дерева, у которого ландшафт функции приспособленности не настолько сложен, чтобы потребовалось включение в работу сразу всех перечисленных выше факторов.

Таблица 2 – Влияние величины вероятности инверсии на количество холостых гипотез

		Вероятность инверсии				
		0.1	0.05	0.01	0.005	0.001
Число холостых гипотез	С аутбридингом и с побитовой мутацией	24	21	5	6	12
	Без аутбридинга и с побитовой мутацией	9	13	14	15	10
	Без аутбридинга и без побитовой мутации	12	11	12	4	9

Для исследуемого дерева [2,3] в том случае, когда при выполнении кроссовера отключен аутбридинг и к тому же не выполняется побитовая мутация, т.е. не работают два мощных фактора предотвращения преждевременной сходимости процесса поиска, а работает только оператор инверсии, получена кривая, дающая наилучшие результаты.

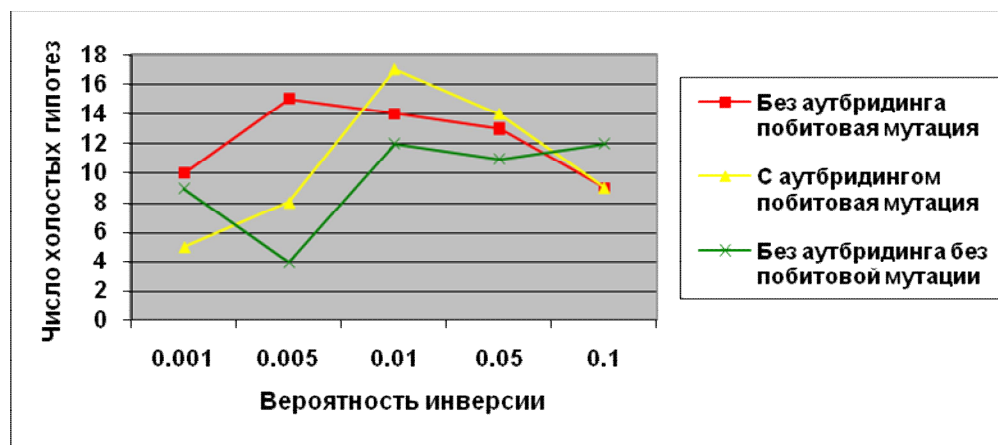


Рисунок 10 – Диаграмма влияния величины вероятности инверсии на эффективность поиска при различных сочетаниях факторов, предотвращающих преждевременную сходимость процесса

Литература

1. Частикова В.А. Оптимизация процессов поиска решений в интеллектуальных системах обработки экспертной информации на основе генетических алгоритмов. Автореферат диссертации на соискание ученой степени кандидата технических наук. – Краснодар: Изд-во КубГТУ, 2005.
2. Частикова В.А. Исследование основных параметров генетического алгоритма метода генетических схем в интеллектуальных системах, основанных на знаниях / В.А. Частикова // Научный журнал КубГАУ [Электронный ресурс]. – Краснодар: КубГАУ, 2011. - № 69(5). – Шифр Информрегистра: 0421100012/0162. – Режим доступа: <http://ej.kubagro.ru/2011/05/pdf/32.pdf>.
3. Симанков В.С., Частикова В.А. Генетический поиск решений в экспертных системах. Монография. - Краснодар: Просвещение-Юг, 2008.
4. Коломиец Т.В., Малыхина М.П. Формирование базы знаний экспертной системы диагностики СУБД. - Известия высших учебных заведений. Северо-Кавказский регион. Серия: Технические науки. 2007. № 3. С. 5-6.
5. Занин Д.Е., Частиков А.П. Эффективность решения задач ранжировки в информационно-поисковых системах на основе динамических нейронных сетей Хопфилда. - Известия высших учебных заведений. Северо-Кавказский регион. Серия: Технические науки. 2008. № 6. С. 62-65.