

УДК 004.023

В.С. Мараев, Е.А. Беззубенко, Д.А. Черкашин, А.С. Михалев
**ИССЛЕДОВАНИЕ ЭФФЕКТИВНОСТИ ГЕНЕТИЧЕСКОГО
АЛГОРИТМА С РАЗЛИЧНЫМИ МЕТОДАМИ СЕЛЕКЦИИ,
ТИПАМИ КРОССОВЕРА И СТРАТЕГИЯМИ ФОРМИРОВАНИЯ
ПОКОЛЕНИЙ ПРИ ПОИСКЕ ЭКСТРЕМУМОВ ФУНКЦИЙ**

*Сибирский Федеральный Университет,
Институт Космических и Информационных технологий*

В данной статье проводится анализ материалов по генетическим алгоритмам. Рассмотрены основные идеи и принципы, лежащие в основе работы генетических алгоритмов. Детально проанализированы базовые этапы работы классического генетического алгоритма. Выполнен обзор наиболее часто встречаемых методов селекции (рулеточный и турнирный), типов кроссовера (одноточечный и равномерный) и стратегий формирования поколений (классическая и элитарная).

На тестовых функциях проведено исследование генетического алгоритма с различными методами селекции, типами кроссовера и стратегиями формирования поколений для поиска глобального минимума. Для каждого вида алгоритма найдена оценка вероятности нахождения истинного решения. Полученные результаты экспериментов тщательно проанализированы. Выявлены достоинства и недостатки различных методов селекции, типов кроссовера, стратегий формирования поколений. Изложены рекомендации по целесообразности применения генетических алгоритмов в различных ситуациях. Определены возможные направления дальнейших исследований.

Ключевые слова: генетический алгоритм, глобальный экстремум, популяция, поколение, селекция, кроссовер, функция Де Джонга, функция Розенброка, функция Растргина.

Генетические алгоритмы (ГА) - это универсальные оптимизационные методы, предложенные американским ученым Джоном Холландом в 1975 году [1]. Алгоритмы основаны на идее эволюции с помощью естественного отбора. Помимо максимально быстрого нахождения экстремума, к сильным сторонам генетических алгоритмов относят и поиск «глобального» экстремума. При оптимизации функции с множеством локальных экстремумов, в отличие от градиентного метода, генетические алгоритмы, в большинстве случаев, позволяют найти «глобальный» минимум, с достаточно быстро обходя локальные.

Алгоритм. На Рисунке 1 изображена схема работы любого генетического алгоритма.

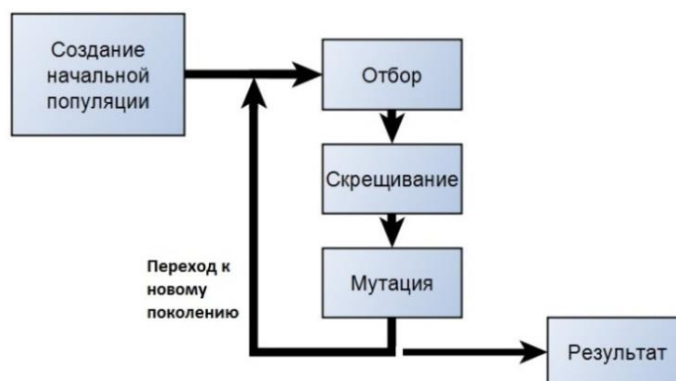


Рисунок 1 - Схема работы генетического алгоритма

Генетические алгоритмы работают с совокупностью особей – популяцией, где каждая особь представляет возможное решение какой-либо задачи. Первая популяция генерируется случайным образом.

Каждое поколение популяция оценивается мерой ее «приспособленности». Наиболее приспособленные особи «воспроизводят» потомство с помощью «перекрестного скрещивания» с другими особями популяции. Благодаря этому создаётся новое поколение, в особях которого, проявляются параметры, полученные от родителей. Особи с малой мерой «приспособленности» смогут воспроизвести потомков с меньшей вероятностью, тем самым, все их неподходящие параметры постепенно исчезнут в процессе эволюции. [2].

Появление новых особей осуществляется не только с помощью скрещивания, но и с помощью мутации. Мутация – случайное изменение генов (рис.2). Каждый бит каждой особи популяции с вероятностью p инвертируется. Эта вероятность обычно достаточно мала.



Рисунок 2 - Мутация

Позже, процесс повторяется. Такой процесс эволюции может продолжаться до бесконечности. Критерием останова может служить заданное количество поколений или схождение популяции.

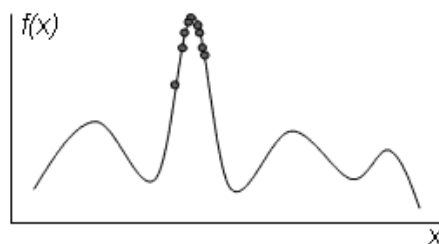


Рисунок 3 - Пример схождения на графике

Схождение – это особое состояние популяции, в котором все гены практически идентичны и пребывают в области определенного экстремума (Рисунок 3). В данной ситуации любое скрещивание минимально повлияет на изменение популяции. Мутация может позволить особи покинуть данную область, однако, такие особи часто имеют низкую приспособленность, особенно если данный экстремум является глобальным. Тем самым, схождение популяции чаще всего свидетельствует о том, что было найдено наиболее оптимальное или близкое к нему решение.

Таким образом, ответом на поставленную задачу будет итоговый набор параметров наилучшей особи последнего поколения.

Размер популяции. Для того, чтобы получить хорошие результаты, необходимо правильно выбрать размер популяции. Действительно, если популяция мала, то при заданном ограничении количества итераций алгоритма создастся большее количество поколений, но скорее всего популяция сойдется на каком-либо локальном экстремуме. Слишком большая популяция скорее всего найдёт решение, но сделать максимально точным его не успеет из за слишком малого количества поколений.

Элитарная стратегия формирования нового поколения. Элитарная стратегия заключается в защите лучших хромосом на некоторых итерациях. Если рассматривать классический алгоритм, то в нем далеко не всегда гены лучших особей наследуются новым поколением. Другими словами, в новой популяции особей не всегда будет хромосома с самыми высокими показателями приспособленности из прошлого поколения. Элитарная система помогает избежать потери таких хромосом. Тем самым, в новой популяции гарантированно останутся лучшие особи предыдущей популяции.

Методы селекции. Чаще всего селекция происходит по принципу рулетки, при этом площадь сегмента колеса рулетки, которое сопоставляется с конкретной особью, пропорциональна величине её приспособленности. Затем рулетка запускается N раз и особь из сектора, на

котором остановилась рулетка, выбирается, однако с рулетки не исчезает. Несмотря на случайный характер процедуры селекции, особи выбираются пропорционально их приспособленности. Недостаток данного метода заключается в том, что особи с очень малой приспособленностью слишком быстро исключаются из популяции, тем самым, генетический алгоритм может сойтись слишком рано. В связи с вышесказанным, были разработаны и успешно используются альтернативные способы селекции. Ярким примером является **турнирная селекция**. При турнирной селекции все особи популяции делятся на группы, затем из каждой группы выбирается особь с лучшей приспособленностью. Допускается изменение размера групп, на которые происходит деление популяции. Теоретически, турнирный метод должен действовать эффективнее, чем метод рулетки.

Типы кроссовера. Существует множество типов кроссовера (скрещивания). Остановимся на следующих двух: одноточечный и равномерный.

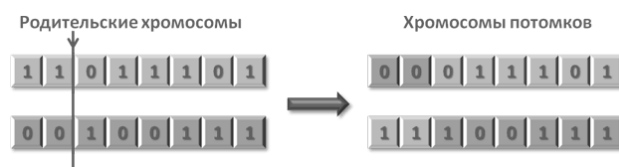


Рисунок 4 - Скрещивание

Одноточечное скрещивание – случайным образом определяется точка разрыва – участок, расположенный между двумя битами. Обе родительские структуры разделяются на два сегмента по этой точке. После чего, соответствующие сегменты двух родителей соединяются, в результате получаются два генотипа потомков (Рисунок 4) [3].

Равномерное скрещивание реализует наследование детьми каждого из генов родителей с некоторой вероятностью. С данным типом скрещивания сложнее добиться четкого схождения, то-есть точность ответа меньше чем при одноточечным скрещиванием, но больше вероятность не застрять на локальном экстремуме. Соответственно, он лучше одноточечного скрещивания при исследовании большого промежутка многоэкстремальной функции и хуже при коротком промежутке унимодальной функции, так как точность ответа в этом случае будет ниже.

Протестируем генетический алгоритм с различными типами селекции, кроссовера и стратегиями формирования поколений для нахождения глобального минимума на тестовых функциях.

Функция Де Джонга. Так называемая сферическая или 1 функция Де Джонга. Является одним из самых простых тестов производительности

алгоритмов оптимизации. Функция непрерывная, выпуклая и унимодальная [4].

Функция имеет следующее общее определение:

$$f(x) = \sum_{i=1}^n x_i^2$$

Допустимая область изменения x : $-5.12 < x_i < 5.12, i = \overline{1, n}$. Глобальный минимум: $f(x) = 0, x_i = 0, i = \overline{1, n}$. Локальных минимумов нет. Пространственный вид минимизируемой функции представлен на Рисунке 5.

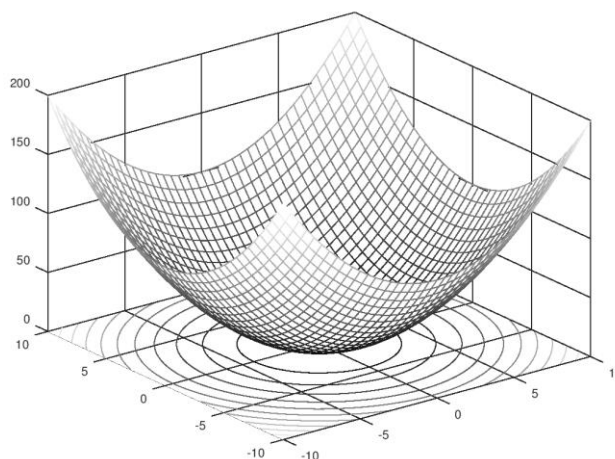


Рисунок 5- График функции Де Джонга при $n = 2$

Генетический алгоритм легко справляется с нахождением минимума данной функции за относительно малое количество итераций.

Как и для любого алгоритма оптимизации, для генетических алгоритмов есть некоторые типы функций, с которыми им работать сложнее, чем с другими. Обычно генетический алгоритм тестируют именно на таких функциях. Примером таких функции являются функции Розенброка и Растригина.

Функция Розенброка. Долина Розенброка - классическая задача оптимизации, также известная как «банановая функция» или 2 функция Де Джонга. Считается, что поиск глобального минимума для данной функции является нетривиальной задачей [5]. Функция имеет большое медленно убывающее плато. Глобальный экстремум функции находится внутри параболической сильно вытянутой поверхности. Определение формы поверхности тривиально, однако сходимость к локальному минимуму - трудная задача.

Функция имеет следующее определение:

$$f(x) = \sum_{i=1}^{n-1} [100 \cdot (x_{i+1} - x_i^2)^2 + (1 - x_i)^2]$$

Допустимая область изменения x : $2.048 < x_i < 2.048, i = \overline{1, n}$. Глобальный минимум: $f(x) = 0, x_i = 0, i = \overline{1, n}$. Пространственный вид минимизируемой функции представлен на Рисунке 6.

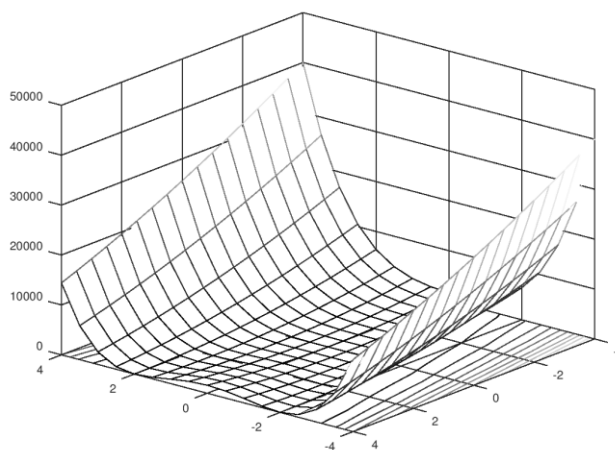


Рисунок 6 - График функции Розенброка при $n = 2$

Функция Растригина. Функция Растригина - невыпуклая функция, используемая для тестирования эффективности алгоритмов оптимизации, типичный пример нелинейной мультимодальной функции [6]. Представляет собой сочетание сферической и синусоидной функций, из-за чего создаётся множество локальных минимумов. Нахождение минимума этой функции является достаточно трудной задачей из-за большой области поиска и большого количества локальных минимумов.

Функция имеет следующее определение:

$$f(x) = 10 \cdot n + \sum_{i=1}^n [x_i^2 - 10 \cdot \cos(2 \cdot \pi \cdot x_i)]$$

Допустимая область изменения x : $-5.12 < x_i < 5.12, i = \overline{1, n}$. Глобальный минимум: $f(x) = 0, x_i = 0, i = \overline{1, n}$. Пространственный вид минимизируемой функции представлен на Рисунке 7.

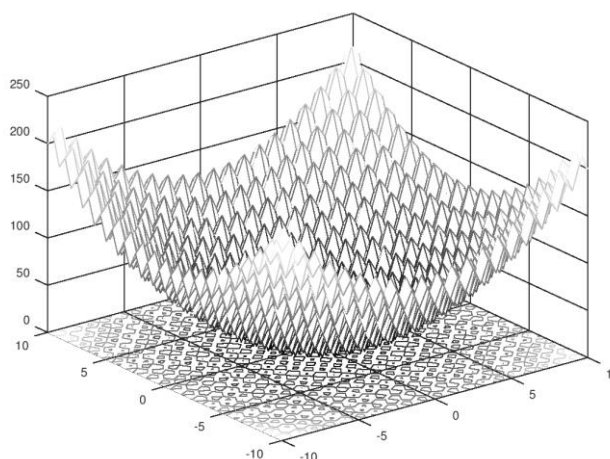


Рисунок 7- График функции Растригина при $n = 2$

Результаты тестирования. Были проведены эксперименты для оценивания возможностей генетического алгоритма с различными типами селекции, кроссовера и стратегиями формирования поколений для нахождения глобального минимума на данных тестовых функциях. Комплексной характеристикой процесса поиска глобального минимума является оценка вероятности нахождения истинного решения $\hat{P}_{ИСТ}$ [7]. Для каждой тестовой задачи было произведено 1000 запусков алгоритма при различных комбинациях параметров. Оценка вероятности нахождения истинного решения $\hat{P}_{ИСТ}$ обычно вычисляется как отношение числа благоприятных исходов («корректных» результатов работы алгоритма) к общему количеству запусков. При этом запуск считается «корректным», если найденное значение целевой функции будет максимально близким к правильному ответу, т.е. если выполняется следующее условие:

$$|f_{ИСТ}^* - f_{НАЙД}^*| \leq \varepsilon.$$

Результаты экспериментов описаны в Таблице 1.

Естественно, что при разном количестве особей и поколений алгоритма, оценки вероятности нахождения истинного решения при разных комбинациях параметров генетического алгоритма будут различны. Обычно, с увеличением числа поколений или числа особей оценка вероятности нахождения истинного решения для каждой комбинации параметров будет стремиться к 1, в свою очередь уменьшение числа особей будет давать обратный результат. Поэтому мы постарались оптимизировать количество особей и поколений так, чтобы каждая

комбинация параметров исследовалась в примерно одинаковых условиях, при этом оценка вероятности ни одной из них не была нулевой или же единичной, чтобы можно было объективно сравнить их характеристики. При всех экспериментах использовались одинаковое число поколений и число особей.

Таблица 1 - Результаты экспериментов

| Характеристика генетического алгоритма | | | Оценка вероятности нахождения истинного решения для разных тестовых функций | | |
|--|----------------|----------------------------------|---|--------------------|-------------------|
| Метод селекции | Тип кроссовера | Стратегия формирования поколения | Функция Де Джонга | Функция Розенброка | Функция Растргина |
| Рулетка | Одноточечный | Классическая | 0.91 | 0.87 | 0.85 |
| Рулетка | Одноточечный | Элитарная | 0.93 | 0.89 | 0.87 |
| Рулетка | Равномерный | Классическая | 0.73 | 0.7 | 0.69 |
| Рулетка | Равномерный | Элитарная | 0.76 | 0.74 | 0.71 |
| Турнир | Одноточечный | Классическая | 0.94 | 0.91 | 0.89 |
| Турнир | Одноточечный | Элитарная | 0.97 | 0.93 | 0.9 |
| Турнир | Равномерный | Классическая | 0.76 | 0.74 | 0.71 |
| Турнир | Равномерный | Элитарная | 0.78 | 0.75 | 0.73 |

Анализ результатов. Как и предполагалось, элитарная стратегия формирования нового поколения оказалась лучше классической, а турнирный метод селекции лучше рулеточного при всех исследуемых функциях. Причины такого исхода уже были описаны ранее. В свою очередь, сравнение типов кроссовера показало, что равномерный тип ведет себя значительно хуже одноточечного. Скорее всего, это связано с тем, что он не слишком оптимален при маленькой допустимой области исследования функций, и если бы мы увеличили интервал, то, вполне вероятно, что он бы сравнялся по качеству с одноточечным кроссовером, о чём уже предполагалось ранее.

Оценка вероятности нахождения истинного решения для функции Де Джонга больше, чем для функции Розенброка, а та, в свою очередь, больше, чем для функции Растргина. Данное замечание справедливо почти для любых параметров генетического алгоритма. Это обусловлено

тем, что функция Растригина сложнее функции Розенброка, а та, в свою очередь, сложнее функции Де Джонга, о чём говорилось ранее.

Как мы видим, наилучшим генетическим алгоритмом в данном исследовании оказался алгоритм с турнирным методом селекции, одноточечным кроссовером и элитарной стратегией формирования нового поколения, а худшим - алгоритм с рулеточным методом селекции, равномерным кроссовером и классической стратегией формирования нового поколения. Это, конечно, не значит, что данные параметры алгоритма будут лучшими или худшими абсолютно всегда, но в большинстве случаев данная зависимость вероятности нахождения истинного решения от данных параметров генетических алгоритмов будет выполняться.

Таким образом, с помощью экспериментов была исследована зависимость оценки вероятности нахождения истинного решения от различных методов селекции, типов кроссовера и стратегий формирования поколений генетических алгоритмов, а также показано как генетический алгоритм с различными методами селекции, типами кроссовера и стратегиями формирования поколений может применяться для нахождения глобального экстремума сложных функций.

Заключение. Генетический алгоритм на сегодняшний день применяются при разработке программного обеспечения, в системах искусственного интеллекта [8], искусственных нейронных сетях [9], для настройки моделей Data Mining [10], а также для построения аналитических систем [11-13]. За счет использования генетических алгоритмов решаются задачи прогнозирования, классификации, поиска оптимальных вариантов, и совершенно незаменимы, когда решение задачи основано на интуиции или опыте.

Генетические алгоритмы предоставляют огромные материалы для исследований за счет большого количества модификаций и параметров. Меняя параметры селекции, кроссовера, стратегии выбора поколений и других этапов алгоритма, можно получать совершенно разные виды генетических алгоритмов, каждый из которых достоин отдельного исследования. После досконального изучения всех видов генетических алгоритмов, исходя из исследований, мы сможем применять наиболее оптимальные виды генетических алгоритмов с определёнными индивидуальными модификациями и параметрами для решения каждой отдельной задачи.

ЛИТЕРАТУРА

1. John R. Koza, Genetic Programming: On the Programming of Computers by Means of Natural Selection - MIT Press, 1992. Vol. 2, No. 2, pp.78-84.
2. Darrel Whitley, A Genetic Algorithm Tutorial – Statistics and Computing, 1994. Vol. 4, pp. 65-85.
3. Herrera F., Losano M., Sanches A.M., Hybrid Crossover Operators for Real-Coded Genetic Algorithms: An Experimental Study.
4. Yang X.-S., Deb S., Engineering optimization by cuckoo search. - Int. J. Math. Modelling Num. Optimisation, 2010. Vol. 1, No. 4, pp. 330 - 343.
5. Rosenbrock H.H., An automatic method for finding the greatest or least value of a function». - The Computer Journal 3, 1960. pp. 175–184.
6. Rastrigin L. A., Systems of Extremal Control - Nauka, Moscow, 1974.
7. Mikhalev A.S., Rouban A.I. Global optimization on set of mixed variables: continuous and discrete with unordered possible values // IOP Conf. Series: Materials Science and Engineering, 2016, Vol. 22, Issue 1: 19th International Scientific Conference Reshetnev Readings 2015.
8. Хант Э. Искусственный интеллект / Пер. с англ. – М.: Изд-во «МИР», 1978. – 281 с.
9. Хайкин С. Нейронные сети: полный курс, 2-е изд. / Пер. с англ. – М.: Изд. дом «Вильямс», 2006. – 1104 с.
10. Вахрушева М.Ю., Глебов М.П. Применение технологии Data Mining в решении демографических проблем// Труды Братского государственного университета. Серия: Экономика и управление. 2013. Т. 1. С. 255-258.
11. Харитоновна П.В. Применение IT-технологий при принятии управленческих решений в малом и среднем бизнесе / Труды Братского государственного университета. Серия: Экономика и управление. 2015. Т. 1. С. 266-269.
12. Евдокимов И.В. Кадровое обеспечение внедрения SCADA-систем на предприятиях//Труды Братского государственного университета. Серия: Экономика и управление. 2005. Т. 1. С. 116-119.
13. Вахрушева М.Ю., Евдокимов И.В. Разработка программного обеспечения аналитических информационных систем//Труды Братского государственного университета. Серия: Экономика и управление. 2014. Т. 1. № 1. С. 196-199.

V.S. Maraev, E.A. Bezzubenko, D.A. Cherkashin, A.S. Mihalev
**RESEARCH OF THE EFFICIENCY OF THE GENETIC ALGORITHM
WITH VARIOUS METHODS OF SELECTION, CROSSOVER TYPES
AND STRATEGIES OF FORMATION OF GENERATIONS IN
SEARCHING OF EXTREMUMS OF TEST FUNCTIONS**

*Siberian Federal University,
The Institute of Space and Information Technology*

In this article the analysis of materials on genetic algorithms is carried out. The main ideas and principles underlying work of genetic algorithms are considered. The basic stages of the classical genetic algorithm work are analyzed in detail. The review of the most common methods of selection (roulette and tournament), types of crossover (single-point and uniform) and strategies of formation of generations (classical and elitist) is executed.

On test functions the research of genetic algorithm with different methods of selection, types of crossover and strategy of formation of generations is carried out. For each type of algorithm, an estimate of the probability of finding a true solution is given. The received results of the experiments are carefully analyzed. The advantages and disadvantages of different methods of selection, types of crossover, strategy of formation of generations are revealed. The recommendations on the expediency of using genetic algorithms in various situations are stated. The possible directions for further research are defined.

Keywords: genetic algorithm, global extremum, population, generation, selection, crossover, De Jong's function, Rosenbrock's function, Rastrigin's function.

REFERENCES

1. John R. Koza, Genetic Programming: On the Programming of Computers by Means of Natural Selection - MIT Press, 1992. Vol. 2, No. 2, pp.78-84.
2. Darrel Whitley, A Genetic Algorithm Tutorial – Statistics and Computing, 1994. Vol. 4, pp. 65-85.
3. Herrera F., Losano M., Sanches A.M., Hybrid Crossover Operators for Real-Coded Genetic Algorithms: An Experimental Study.
4. Yang X.-S., Deb S., Engineering optimization by cuckoo search. - Int. J. Math. Modelling Num. Optimization, 2010. Vol. 1, No. 4, pp. 330 - 343.
5. Rosenbrock H.H., An automatic method for finding the greatest or least value of a function». - The Computer Journal 3, 1960. pp. 175–184.
6. Rastrigin L. A., Systems of Extremal Control - Nauka, Moscow, 1974.
7. Mikhalev A.S., Rouban A.I. Global optimization on set of mixed variables: continuous and discrete with unordered possible values // IOP Conf. Series: Materials Science and Engineering, 2016, Vol. 22, Issue 1: 19th International Scientific Conference Reshetnev Readings 2015.
8. Khant E. Iskusstvennyy intellekt / Per. s angl. – M.: Izd-vo «MIR», 1978. – p. 281

9. Khaykin S. Neyronnye seti: polnyy kurs, 2-e izd. / Per. s angl. – M.: Izd. dom «Vil'yams», 2006. – p.1104.
10. Vakhrusheva M.Yu., Glebov M.P. Primenenie tekhnologii Data Mining v reshenii demograficheskikh problem// Trudy Bratskogo gosudarstvennogo universiteta. Seriya: Ekonomika i upravlenie. 2013. Vol. 1. pp. 255-258.
11. Kharitonova P.V. Primenenie IT-tekhnologiy pri prinyatii upravlencheskikh resheniy v malom i srednem biznese / Trudy Bratskogo gosudarstvennogo universiteta. Seriya: Ekonomika i upravlenie. 2015. Vol. 1. pp. 266-269.
12. Evdokimov I.V. Kadrovoe obespechenie vnedreniya SCADA-sistem na predpriyatiyakh//Trudy Bratskogo gosudarstvennogo universiteta. Seriya: Ekonomika i upravlenie. 2005. Vol. 1. pp. 116-119.
13. Vakhrusheva M.Yu., Evdokimov I.V. Razrabotka programmogo obespecheniya analiticheskikh informatsionnykh sistem//Trudy Bratskogo gosudarstvennogo universiteta. Seriya: Ekonomika i upravlenie. 2014. Vol. 1. No. 1. pp. 196-199.