

Российская ассоциация искусственного интеллекта
Академия наук Республики Татарстан
Казанский федеральный университет
Институт проблем управления РАН
Институт системного анализа РАН
Институт вычислительной математики и информационных технологий КФУ
Высшая школа информационных технологий и информационных систем КФУ
Научно-исследовательский институт "Прикладная семиотика" АН РТ

КИИ-2014

ЧЕТЫРНАДЦАТАЯ НАЦИОНАЛЬНАЯ КОНФЕРЕНЦИЯ ПО ИСКУССТВЕННОМУ ИНТЕЛЛЕКТУ С МЕЖДУНАРОДНЫМ УЧАСТИЕМ

**Программа конференции
24-27 сентября 2014 г.**

**Казань
Россия**

<p>оптимального расписания передачи информации в форме знаний для сетей с древовидной топологией 6. Колесов А.О., Наместников А.М. (Ульяновск). Интеграция реляционных данных на основе онтологического подхода</p>	<p>пространства признаков с учетом поведения ошибок обобщения 10. Котельников Е.В. (Киров). Повышение быстродействия ДСМ-метода в задачах обработки текстовой информации</p>	<p>позиции модели «ЭТАП» 5. Кириков И.А., Колесников А.В., Листопад С.В. (Калининград). Гибридные интеллектуальные системы: координация и согласованность 6. Кулинич А.А. (Москва). Когнитивная архитектура систем поддержки принятия решений: семиотический подход</p>
<p>16.00 – 16.15 Кофе-брейк</p>		
<p>16.15 – 18.00</p>		
<p>Секция 11 Инженерия знаний Председатели: Т.А.Гаврилова, В.В.Грибова 7. Клещев А.С., Шалфеева Е.А. (Владивосток). Поддержка управления базами знаний при автоматизации интеллектуальной профессиональной деятельности 8. Курбатов С.С Лобзин А.П. . Хахалин Г.К (Москва). Программное обеспечение для лингвистически-ориентированного пополнения онтологии 9. Мошкин В.С., Ярушкина Н.Г. (Ульяновск). Особенности интеграции механизмов логического вывода в онтологическую модель представления знаний с помощью SWRL-правил 10. Сулейманов Д.Ш., Аюпов М.М., Невзорова О.А., Прокопьев Н.А. (Казань). Семантические технологии генерации учебных вопросов</p>	<p>Секция 9 Интеллектуальный анализ данных и машинное обучение Председатели: Д.В.Виноградов, М.А.Михеенкова 11. Кохов В.В. (Москва). Классификация задач определения сходства орграфов 12. Кохов В.В. (Москва). Задачи и методы анализа сходства темпоральных орграфов 13. Михеенкова М.А., Финн В.К. (Москва). Интеллектуальные системы и проблемы поддержки социологических исследований 14. Найденова К.А. (Санкт-Петербург). Итоги и перспективы развития модели классификации на основе хороших диагностических тестов (часть 1) 15. Панов А.И., Швец А.В. (Москва). Эволюционный метод покрытий для составления базы фактов ДСМ-метода 16. Суворова А.В. Хайбуллин Р.Р. Тулупьев А.Л. (Санкт-Петербург). Байесовская сеть доверия как инструмент оценивания интенсивности рискованного поведения респондентов</p>	<p>Секция 6 Интеллектуальные системы поддержки принятия решений Председатели: А.П.Еремеев, А.Б.Петровский 7. Лобанов В.Н., Петровский А.Б. (Москва). Особенности построения иерархической системы критериев для оценки сложного технического комплекса 8. Перфильева И.Г. (Острава, Чехия), Романов А.А. (Ульяновск). Извлечение знаний о группах временных рядов для задач прогнозирования 9. Ройтман Е.Я., Ройзензон Г.В. (Москва). Многокритериальный подход в задаче распределения ресурсов 10. Янковская А.Е., Гедике А.И. (Томск). Интеллектуальная подсистема построения отказоустойчивых безусловных диагностических тестов и принятия решений на их основе</p>
<p>18.00 – 19.00 Вечерняя лекция В.Ф. Хорошевский, И.В. Ефименко</p>		
<p>19.15 – 19.30. Информационное сообщение: А.В. Заболева-Зотова (Москва). О региональных конкурсах РФФИ</p>		

26 сентября, пятница

<p>10.00 – 10.45 Пленарный доклад: О.П.Кузнецов, Л.Ю.Жилиякова. Сложные сети и когнитивные науки</p>		
<p>10.45 – 11.30 Пленарный доклад: В.Е.Павловский. Интеллектуальная робототехника: современное состояние проблемы</p>		
<p>11.30 – 11.45 Кофе-брейк</p>		
<p>11.45 – 13.15</p>		
<p>Секция 5 Многоагентные системы Председатели: В.И.Городецкий, В.Б.Тарасов 1. Базенков Н.И. (Москва). Теоретико-игровые алгоритмы формирования беспроводной ad hoc сети</p>	<p>Секция 12 Выставка программных продуктов Председатели: В.Ф.Хорошевский, И.А.Тихомиров 1. Берман А.Ф., Малтугуева Г.С., Юрин А.Ю. (Иркутск).</p>	<p>Секция 1 Моделирование рассуждений Председатели: В.Н.Вагин, О.М.Аншаков 1. Вагин В.Н., Моросин О.Л. (Москва). Обзор методов нахождения степеней обоснования в системах аргументации</p>

<p>2. Дюндюков В.С., Тарасов В.Б. (Москва) Многосторонние ресурсно-целевые сети в моделировании взаимодействий между когнитивными агентами</p> <p>3. Куливец С.Г. (Москва). Повторяющаяся игра с линейными функциями полезности на когнитивной карте с ограничениями и без ограничений на ресурсы</p> <p>4. Пантелеев М.Г. (Санкт-Петербург). Формальная модель опережающего итеративного планирования действий интеллектуальных агентов реального времени</p>	<p>Интеллектуальная система поддержки принятия решений при обосновании выбора конструкционных материалов</p> <p>2. Загорулько Г.Б. (Новосибирск). Создание портала знаний для информационно-аналитической поддержки разработчиков СППР</p> <p>3. Загорулько Ю.А., Боровикова О.И., Сидорова Е.А., Ахмадеева И.Р. (Новосибирск). Подход к автоматизации сбора информации для тематических интеллектуальных научных интернет-ресурсов</p> <p>4. Иванов В.К., Палюх Б.В., Мескин П.И. (Тверь). Прототип программной реализации генетического алгоритма для документального поиска</p>	<p>2. Виньков М.М., Фоминых И.Б. (Москва). Моделирование поведения интеллектуального агента при непредвиденном увеличении длительности его дедуктивного цикла в условиях жёстких временных ограничений</p> <p>3. Захаров А.С. (Смоленск). Особенности построения нечётких байесовских сетей доверия для моделирования темпоральных рассуждений</p> <p>4. Зуенко А.А. (Апатиты). Обработка специальных видов ограничений при решении задач удовлетворения ограничений с помощью алгебры кортежей</p> <p>5. Зуенко А.А. (Апатиты). Распространение ограничений и эвристический поиск с применением матричного представления конечных предикатов</p>
<p>13.15 – 14.00 Обед</p>		
<p>14.00 – 15.30</p>		
<p>Секция 5 Многоагентные системы Председатели: В.И.Городецкий, В.Б.Тарасов</p> <p>5. Тельнов Ю.Ф. (Москва) Модель многоагентной системы реализации информационно-образовательного пространства</p> <p>6. Трёмбач В.М. (Москва). Многоагентная система для решения задач целенаправленного поведения</p> <p>7. Федотов А.М., Ломакин С. Г. (Новосибирск). Модель самоорганизации в агентных системах с передачей сообщений</p>	<p>Секция 12 Выставка программных продуктов Председатели: В.Ф.Хорошевский, И.А.Тихомиров</p> <p>5. Котенко И.В., Саенко И.Б. (Санкт-Петербург). Интеллектуальная система мониторинга и управления инцидентами кибербезопасности</p> <p>6. Субхангулов Р.А., Филиппов А.А. (Ульяновск). Онтологически-ориентированная система интеллектуального анализа технических документов</p> <p>7. Усталов Д.А. (Екатеринбург). Автоматическая генерация заданий для краудсорсингового тезауруса</p>	<p>Секция 10 Нечеткие модели и мягкие вычисления Председатели: А.Н.Аверкин, Н.Г.Ярушкина</p> <p>1. Афанасьева Т.В., Заварзин Д.В. (Ульяновск). Алгоритм поиска аномалий в процессах на основе нечетких тенденций временных рядов</p> <p>2. Зернов М.М. (Смоленск). Модель нечеткого сценария развития временного ряда</p> <p>3. Ковалев С.М. (Ростов-на Дону). Нечетко-стохастический вывод в интеллектуальных системах обработки первичной информации</p> <p>4. Куценко Д.А., Поляков В.М., Синюк В.Г. (Белгород). Метод вывода на основе нечеткой степени истинности для систем со многими нечеткими входами</p>
<p>15.30 – 15.45 Кофе-брейк</p>		
<p>15.45 – 19.00</p>		
<p style="text-align: center;">СЪЕЗД РАИИ</p>		
<p>20.00 Мероприятие №2 по плану Оргкомитета</p>		

УДК 004.89:004.424.4:519.87

ПРОТОТИП ПРОГРАММНОЙ РЕАЛИЗАЦИИ ГЕНЕТИЧЕСКОГО АЛГОРИТМА ДЛЯ ДОКУМЕНТАЛЬНОГО ПОИСКА¹

В.К. Иванов (*mtivk@tstu.tver.ru*),

Б.В. Палюх (*pboris@tstu.tver.ru*),

П.И. Мескин (*pavel.meskin@gmail.com*)

Тверской государственной технической университет,
Тверь

В статье представлен прототип программной реализации генетического алгоритма, который может использоваться для формирования эффективных поисковых запросов и отбора релевантных результатов при выполнении документального тематического поиска. Рассматриваются целевая функция, основные шаги и параметры алгоритма. Описываются компоненты программного обеспечения. Приведены некоторые результаты предварительных исследований алгоритма.

Введение

Каждый пользователь Интернет с помощью поисковых систем искал в сети какие-либо конкретные факты и, как правило, находил либо точно то, что искал, либо что-то близкое. Результат такого поиска – описание объекта, события или явления с заданными значениями их свойств.

В других случаях мы ищем целые категории скоординированной информации в некотором тематическом сегменте, а не отдельные информационные объекты с заданными характеристиками (представителей этих категорий). Результатом, кроме набора фактов, следует считать сведения о ретроспективе, перспективе, взаимосвязях найденных информационных объектов, о текущих и вероятных трендах.

При выполнении тематического поиска неизбежно возникает ряд вопросов. Как совместно оценить релевантность документов, найденных разными запросами? Является ли ранжирование результатов корректным с позиций ожиданий пользователя? Все ли результаты, соответствующие ожиданиям пользователя, попали в число доступных для

¹ Работа выполнена при финансовой поддержке РФФИ (проект № НК13-07-00342).

непосредственной оценки? Как отфильтровать документы, не относящиеся по сути к искомой тематике? Могут ли быть обнаружены эффективные решения, которые относятся к другим областям применения, но могут быть успешно использованы в данной области? Невозможно однозначно ответить на эти вопросы в рамках тривиальных решений.

Для получения эффективного множества поисковых запросов мы предлагаем использовать подход, основанный на генетических алгоритмах, область применения которых может быть достаточно широкой (см., например, системное изложение в [Курейчик и др., 2006]). Подход предусматривает организацию эволюционного процесса, а) формирующего устойчивую и эффективную популяцию поисковых запросов, б) образующего соответствующий поисковый образ документов или семантического ядра, в) приводящего к получению релевантного искомого множества документов.

Примененные подходы и технологии

Представленная в данной статье разработка, являясь частью проекта "Интеллектуальная распределенная система информационной поддержки инноваций в науке и образовании" [Палух и др., 2013], касается решения задач фильтрации данных [Napani et al., 2001], [Маннинг и др., 2011] на основе непосредственного использованием контента. Для нас представляли существенный интерес исследования видов поиска в Web [Broder, 2002] и различных вариантов альтернативных техник поиска, таких как использование паттернов [Zhu et al., 2013] с возможностью специфицировать информационные потребности в виде набора релевантных документов (а не набора ключевых слов) и оптимальное совместное использование нескольких информационных источников для получения набора запросов [Wu et al., 2011] и [Bendersky et al., 2012].

Отметим, что для различных программных платформ разработано довольно большое количество реализаций генетических алгоритмов. Есть как не потерявшие своей актуальности разработки [Wall, 1996], так и современные реализации, например, GeneHunter (<http://www.wardsystems.com>) или Genetic Algorithm Framework for .Net (<http://johnnewcombe.net/gaf>). Некоторые авторы делают попытки унификации и стандартизации подходов к разработкам [Сергиенко и др., 2010], а другие ориентируются на специальное применение генетических алгоритмов [Подлазова, 2008].

При разработке обсуждаемого в статье алгоритма было принято решение об оригинальной реализации основных генетических операций. Причины следующие:

1. Особенности вычисления значений целевой функции.

2. Представление поисковых запросов как хромосом, гены которых (ключевые слова) выражены значениями номинальной шкалы.

3. Специфические интерпретации базовых генетических операций – обмен понятийными элементами и использование синонимии.

4. Необходимость исследований, анализа и выполнения потенциальных улучшающих модификаций алгоритма.

5. Планируемое использование генетического алгоритма в интеллектуальном поисковом приложении для мобильных устройств.

Целевая функция

Для каждого i -го результата запроса целевая функция (пригодность) может быть определена как $w_i = f(p, r, s)$, где

p – средний номер позиции адреса документа в списке первых P результатов выполненных поисковых запросов (учитываются только те списки результатов, где данный адрес присутствует).

r – количество появлений адреса документа в результатах выполнения N поисковых запросов. Отметим, что $r \leq N$ и $r = N$, если адрес документа появился в результатах выполнения всех запросов.

s – семантическая близость текста найденного документа (по крайней мере заголовка и сниппета) и поискового образа документов (исходного множества ключевых слов). В качестве альтернативы используются эталонные тексты, адаптивно формируемые в ходе выполнения алгоритма. Для вычисления s использована модификация модели векторного пространства документов, в которой каждый документ интерпретируется как вектор $\vec{v}(d) = (w_{1,d}, w_{2,d}, \dots, w_{M,d})$, где $w_{i,d} = tf_{i,d} * idf_{i,d}$. Здесь $tf_{i,d}$ – частота использования термина в документе, $idf_{i,d} = \log(P + 1/P_i)$ где P – общее число документов в SERP), P_i – число документов, содержащих данный термин, M – число терминов в P документах. Близость текстов s интерпретирована как косинусная мера близости $s(d_1, d_2)$.

Значения p, r, s нормированы на диапазон от 0 до 1. Ранг каждого результата запроса вычисляется как средневзвешенное значение $w_i = f_5 * p + f_6 * r + f_7 * s$, где f_5, f_6, f_7 – весовые коэффициенты, являющиеся параметрами алгоритма, $f_5 + f_6 + f_7 = 1$.

Значение целевой функции для запроса \bar{w} – средний ранг результатов этого запроса, а значение целевой функции для популяции запросов \bar{W} – средний ранг запросов этой популяции.

Описание алгоритма

На рис. 1 приведена схема генетического алгоритма формирования эффективного множества запросов. Ниже кратко описаны основные шаги алгоритма; более подробное описание см. в [Иванов, 2013].

1. Подготовка ключевых слов. Множество ключевых слов $K = \{k_j\}$, $Q_0 = \{k\}$, формирует поисковый образ множества документов заданной тематики.

2. Выбор поисковой системы. Используется любая поисковая система (или несколько одновременно), имеющая API: Bing, Google и др.

3. Формирование исходной популяции запросов. Каждый поисковый запрос есть совокупность ключевых слов. Исходная популяция из N запросов – это множество $Q = \{q_i\}$, $|Q| = N$, где $q_i = \{k_1, k_2, \dots, k_m\}$ – случайно образованная комбинация ключевых понятий из K , $k_j \in K$ и для любой пары (k_r, k_s) из K справедливо $k_r \neq k_s$.

4. Выполнение запросов популяции. Результатом являются возможно пересекающиеся множества дескрипторов найденных документов (заголовков, описание, адрес текста).

5. Вычисление целевой функции.

6. Селекция лучших запросов. Следующее поколение формируется теми запросами, у которых $\bar{w} \geq \bar{W}$, где порог \bar{W} может быть средним значением или медианой по текущей популяции.

7. Выбор родительских пар запросов. Используется генотипный аутбридинг. Расстояние между родителями вычисляется как $\Delta \bar{w} = \bar{w}_1 - \bar{w}_2$.

8. Скрещивание запросов. Реализуется одноточечным кроссинговером – обмен ключевыми словами или их группами между запросами.

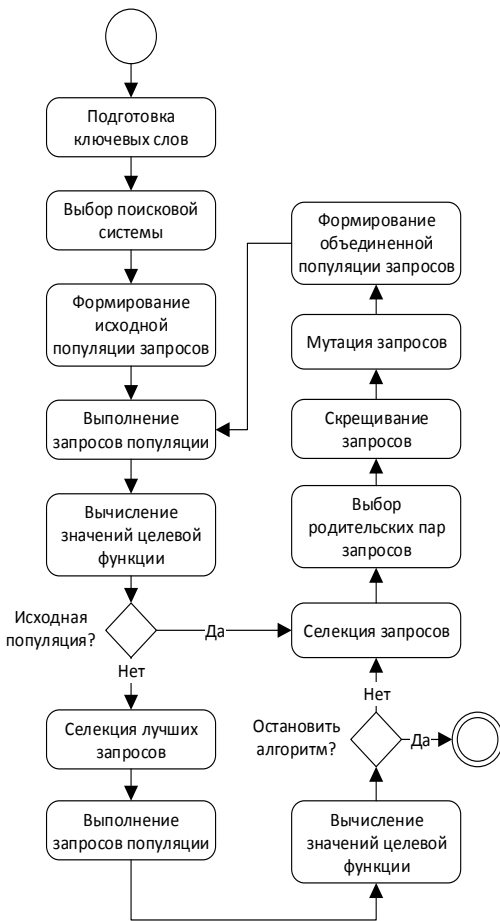


Рис.1. Генетический алгоритм формирования эффективного множества запросов

Особенность реализации – ключевое слово k_i запроса-родителя замещается синонимом слова из запроса-родителя $k_{st} \in S_i$, где S_i – множество синонимов слова k_i (генерируется существенно больше потомков при сохранении семантики родителей).

9. Мутация запросов. Наиболее адекватная операция мутации в нашем подходе – замена случайно выбранного слова запроса на его синоним. Невозможно применение присоединения или вставки нового термина, удаления термина, обмена местами терминов.

10. Формирование объединенной популяции. Используется элитарный отбор. Лучшие решения не теряются.

11. Остановка алгоритма. Общее условие остановки – стабильность популяции. Например, достижение среднееквadraticным отклонением значений целевой функции пороговой величины. Есть варианты использования заданного или предельного числа проходов алгоритма.

Параметры алгоритма

Основные параметры: g_1 – поисковая система (например, Bing или Google); g_2 – количество запросов в генерируемых популяциях; g_3 – количество ключевых слов в каждом генерируемом запросе; g_4 – исходный набор ключевых слов и понятий.

Параметры целевой функции: f_1 – количество результатов поиска, возвращаемых запросом; f_2 – количество результатов поиска, возвращаемых после ранжирования результатов популяции; f_3 – количество результатов поиска, возвращаемых после ранжирования результатов всех популяций; f_4 – коэффициент, учитывающий расположение документов на одном сервере; f_5, f_6, f_7 – веса аргументов p, r и s соответственно при расчете ранга результата поиска; f_8 – способ вычисления целевой функции для групп результатов.

Параметры скрещивания и мутации: c_1 – множитель, используемый для вычисления критерия отбора запросов-родителей: $\bar{w} > c_1 * \overline{Ww} > c_1 * \overline{W}$; m_1 – вероятность мутации запроса.

Параметры завершения алгоритма: e_1 – заданное число проходов алгоритма; e_2 – предел среднееквadraticного отклонения целевой функции $\sigma < e_2$; e_3 – предельное число проходов алгоритма.

Реализация алгоритма

Основные функциональные компоненты приложения на платформе .NET Framework, реализующего алгоритм, следующие:

1. Пользовательский интерфейс на базе Windows Form.
2. Основная библиотека алгоритма GAF. Объектная модель реализует следующие основные классы: *GAF* – свойства, операции, входные и выходные данные алгоритма; *Options* – параметры алгоритма; *Population* –

свойства и операции популяции; *Individual* – свойства и операции запроса популяции. Класс *Individual* имеет клоны *SelectedParents* (запросы-родители), *SelectedParentsPairs* (пары запросов-родителей), *Children* (запросы-потомки), *MutedChildren* (мутировавшие запросы); *SearchResult* – свойства и операции элемента результатов запроса; *TextSimilarity* – определяет параметры вычисления похожести текстов.

3. Модуль морфологического анализа и лемматизации запросов. Используется технология phpMorphy с бинарными словарями (<http://phpmorphy.sourceforge.net>).

4. Модуль семантического анализа сходства текстов. Используется экспериментальная платформа, описанная в [Иванов, 2012].

5. Модули поиска, обеспечивающие унификацию коллекций результатов на базе Bing Search API и Google Custom Search API.

6. Модуль управления базой данных. Обеспечивает доступ к используемым словарям и сохранение результатов работы алгоритма.

7. Модуль управления метаданными (разработан на базе XML-схем). Обеспечивает обработку параметров.

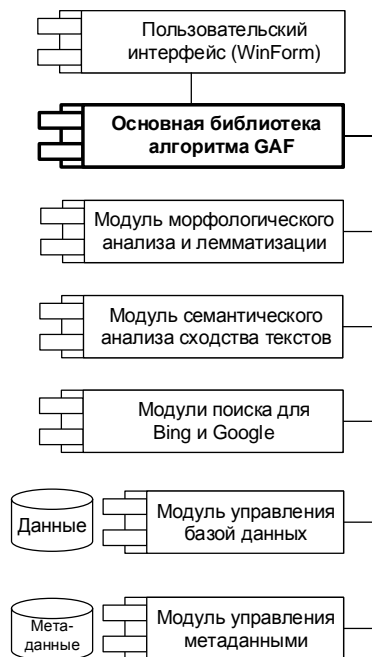


Рис.2. Основные компоненты приложения, реализующего генетический алгоритм

Предварительное исследование алгоритма

В экспериментах использовалась программная реализация алгоритма, описанная выше, и следующий исходный набор значений параметров: $g_1 = \text{Bing}$, $g_2 = 15$, $g_3 = 3$, $f_1 = 10$, $f_2 = 50$, $f_3 = 50$, $f_4 = 0,75$, $f_5 = 0,33$, $f_6 = 0,33$, $f_7 = 0,34$, $f_8 = \text{Average}$, $c_1 = 0$, $m_1 = 1$, $e_1 = 10$, $e_2 = 0$, $e_3 = 50$. В исходном наборе g_4 использовались термины с наибольшими весами $tf_{i,d} * idf_{i,d}$ из корпуса документов – студенческих рефератов.

На рис. 3 показаны графики зависимости \bar{W} от номера популяции при $g_4 = 50$. Видны локальные максимумы \bar{W} и точки начала относительной стабилизации \bar{W} (5-6-я популяция).

На рис. 4 показаны графики зависимости \bar{W} от g_3 (показаны числами около графиков). Видно, что увеличение g_3 в целом приводит к улучшению качества популяций.

На рис. 5 показано влияние аргументов целевой функции на ее значение. Наибольшее влияние оказывает p , наименьшее – r .

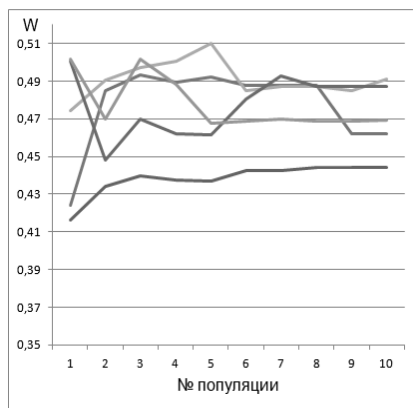


Рис. 3. Изменение \bar{W} ($g_4=50, e_1=10$)

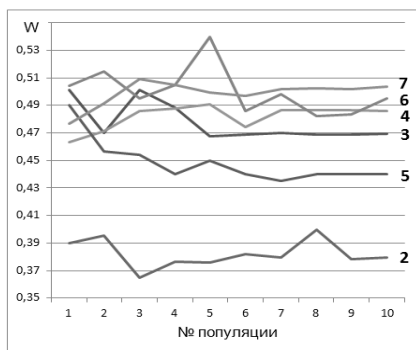


Рис. 4. Изменение \bar{W} ($g_4=50, g_3=2\dots7$)

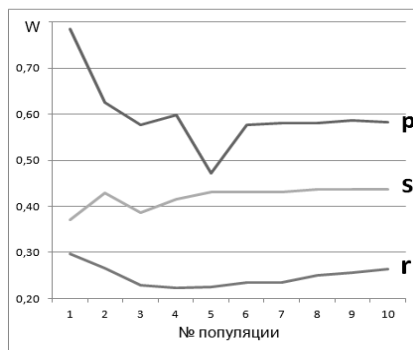


Рис. 5. Влияние аргументов на \bar{W}

Заключение

Генетический алгоритм, рассмотренный в настоящей статье, является одним из элементов программного обеспечения разрабатываемой интеллектуальной системы информационной поддержки инноваций в науке и образовании. В настоящее время проводятся исследования эффективности алгоритма с точки зрения пользователей.

Описанные подходы к организации тематического поиска могут быть успешно применены в ряде областей. Например, обзоры источников научно-технической, коммерческой и социальной информации, поиск коммерчески ценных данных, сбор информации о клиентах, определение

новых областей при бизнес-планировании, конкурентный анализ и разведка, поиск инновационных решений, подбор и экспертиза учебно-методических материалов, анализ конкурсной документации и условий экспертизы проектов, подборка материалов для патентных исследований.

Список литературы

[Иванов, 2012] Иванов В.К. Критерии интегральной оценки электронных документов в системах подготовки принятия решений: статья // Вестник ТвГТУ: Вып. 22. – С. 20-26. – Тверь, 2012.

[Иванов, 2013] Иванов, В.К. Основные шаги генетического алгоритма фильтрации результатов тематического поиска документов : статья // Инновации в науке. – Новосибирск, 2013. – № 25. – С. 8-15.

[Курейчик и др., 2006] Гладков Л.А., Курейчик В.В., Курейчик В.М. Генетические алгоритмы / Под ред. В.М. Курейчика. – 2-е изд., испр. и доп. – М.: Физматлит, 2006. – 320 с.

[Маннинг и др., 2011] Маннинг, Кристофер Д., Рагхаван, Прабхакар, Щютце, Хайнрих. Введение в информационный поиск. : Пер. с англ. – М. : ООО «И.Д. Вильямс», 2011. – 528 с.

[Палюх и др., 2013] Палюх Б.В., Иванов В.К., Сотников А.Н.. Архитектура интеллектуальной системы информационной поддержки инноваций в науке и образовании // Программные продукты и системы. – 2013. – № 4. – С. 197-202.

[Подлазова, 2008] Подлазова, А. В. Генетические алгоритмы на примерах решения задач раскрой // Проблемы управления. – 2008. - №2. – С. 57-63.

[Сергиенко и др., 2010] Генетический алгоритм. Стандарт: Ч. 1: Описание стандартного генетического алгоритма (сГА) / А. Б. Сергиенко, П. В. Галушин, В. В. Бухтояров [и др.]; Сибирский гос. аэрокосмический ун-т им. М. Ф. Решетнева. – Красноярск, 2010. - 384 с.

[Bendersky et al., 2012] Bendersky M., Metzler D., Croft W. B. Effective query formulation with multiple information sources. WSDM'12 Proceedings of the fifth ACM international conference on Web search and data mining, pp. 443-452, DOI: 10.1145/2124295.2124349.

[Broder, 2002] Broder A. A taxonomy of web search, ACM SIGIR Forum Vol. 36, Issue 2, Fall 2002, pp. 3 – 10, DOI: 10.1145/792550.792552.

[Hanani et al., 2001] Hanani U., Shapira B., Shoval P. Information Filtering: Overview of Issues. Research and Systems, User Modeling and User-Adapted Interaction II: pp. 203-259. 2001.

[Wall, 1996] Wall M. Galib: A C++ library of genetic algorithm components. Mechanical Engineering Department, Massachusetts Institute of Technology, 87, 1996.

[Wu et al., 2011] Wu J., Ilyas I., Weddell G. A Study of Ontology-based Query Expansion, Cheriton School of Computer Science, University of Waterloo, Technical Report CS-2011-04, February 09, 2011, p. 38

[Zhu et al., 2013] Zhu M., Xu C., Wu Y.-F.B. IFME: information filtering by multiple examples with under-sampling in a digital library environment. JCDL'13 Proceedings of the 13th ACM/IEEE-CS joint conference on Digital libraries, pp. 107-110, DOI: 10.1145/2467696.2467736.