

# Кластеризация ситуаций при решении прикладных оптимизационных задач (на примерах задач коммивояжёра и восстановления матрицы расстояний)

Б. Ф. Мельников, А. В. Ничипорчук, М. А. Тренина, М. Э. Абрамян

**Аннотация**—В задачах дискретной оптимизации мы применяем алгоритмы, основанные на расширениях метода ветвей и границ. Сами эти расширения заключаются в совместной работе нескольких вспомогательных эвристических алгоритмов, они могут быть отнесены к разным, причём независимым друг от друга, областям искусственного интеллекта. Поэтому актуальность рассматриваемых нами задач обеспечивается как предметными областями, так и алгоритмами.

В статье мы исследуем возможность применения одного из таких вспомогательных алгоритмов – т.н. кластеризации ситуаций. В качестве предметных областей мы рассматриваем две разные задачи дискретной оптимизации: задачу коммивояжёра в её классической постановке (при этом мы отдаём предпочтение исследованию её частных случаев, полученных для псевдогеометрической версии) и задачу восстановления матрицы расстояний ДНК. В результате вычислительных экспериментов мы получили некоторые закономерности, позволяющие создавать улучшенные версии алгоритма ветвей и границ – с помощью подключения к нему эвристик для кластеризации ситуаций.

Результаты, полученные в вычислительных экспериментах, дают обоснование применения кластеризации ситуаций при разработке алгоритмов с помощью метода ветвей и границ. Например, для задачи коммивояжёра такое применение даёт легко наблюдаемые улучшения работы алгоритма – прежде всего для псевдогеометрического варианта.

**Ключевые слова**—эвристические алгоритмы, задачи дискретной оптимизации, метод ветвей и границ, матрица расстояний ДНК, кластеризация ситуаций.

## I. Введение

Настоящая статья является продолжением нескольких наших предыдущих работ, [1], [2], [3], [4] и др.; мы, среди прочего, обобщаем здесь результаты статьи [4]. Мы продолжаем рассмотрение применения *мультиэвристического подхода* к задачам дискретной оптимизации. В нём мы к «обычным» вариантам метода ветвей и границ добавляем несколько вспомогательных эвристических алгоритмов, практически одинаково реализуемых в различных предметных областях. Важно отметить,

Статья получена 1 февраля 2019.

Борис Феликсович Мельников, Российский государственный социальный университет (email: bf-melnikov@yandex.ru).

Анастасия Владимировна Ничипорчук, Российский государственный социальный университет (email: anichiporchuk@gmail.com).

Марина Анатольевна Тренина, Тольяттинский государственный университет (email: trenina.m.a@yandex.ru).

Михаил Эдуардович Абрамян, Южный федеральный университет (email: m-abramyan@yandex.ru).

что наибольший эффект эти вспомогательные эвристики обычно дают *при их одновременном применении*, т.е. в комплексе.

В статье мы рассмотрим следующие вопросы:

- статистическое исследование различных версий ЗКВ средней размерности (раздел II);
- статистическое исследование восстановления матриц ДНК малых размерностей (раздел III);
- краткое описание примера применения метода ветвей и границ в другой задаче – логически связанной с основной задачей, рассматриваемой нами в настоящей статье (раздел IV);
- а также пример малой размерности псевдогеометрической версии задачи коммивояжёра (раздел V).

## II. Краткое статистическое исследование различных версий ЗКВ средней размерности

В этом разделе приведены результаты проведённого нами *статистического* исследования на матрицах средней размерности – для трёх различных *вариантов* задачи коммивояжёра: т.н. геометрической, псевдогеометрической и случайной задач.

Сразу оговоримся про термин «средняя размерность» (мы её понимаем как размерность примерно 60–70, см. далее): ведь, казалось бы, этот термин неудачен, поскольку ещё в конце 1990-х годов появились эвристические алгоритмы, дающие приемлемое псевдо-оптимальное решение для значительно больших размерностей (порядка нескольких миллионов) – см. на эту тему различные публикации в сборнике [5]. Однако все эти алгоритмы применялись *только* к т.н. геометрическому (Евклидову) варианту ЗКВ: даже некоторая его модификация приводит к тому, что мы не можем применять т.н. «алгоритмы луковой шелухи». Поэтому приходится применять совершенно иные подходы – среди которых, по мнению авторов, различные модификации метода ветвей и границ представляют собой наиболее удачный вариант. И, согласно неоднократно проведённым нами вычислительным экспериментам, на современной вычислительной технике методом ветвей и границ (и его простейшими модификациями) для *случайной* ЗКВ можно ожидать успеха (*точного* решения заданного варианта проблемы за реальное время – скажем, за 1 час – с близкой к 1 вероятностью) при размерности задачи 60 и менее. А при размерности 70 и более соответствующая вероятность,

наоборот, будет близка к 0 – поэтому размерность 65 мы и называем «средней».

Результаты проведённых нами вычислительных экспериментов приведены в таблице 1. Мы отвечали только на один вопрос: *насколько часто встречается хотя бы 1 пара одинаковых матриц при применении 30 первых шагов «классического» МВГ.*

При этом генерация входных данных осуществлялась в соответствии с тремя вариантами ЗКВ:

- случайным (в приведённой таблице ниже столбец помечен «случ.»);
- псевдо-геометрическим (два столбца с разными значениями  $\sigma$  эти же значения, 0.2 и 1.4, уже использовались нами в [6] – помечены «псевдо»);
- геометрическим (столбец помечен «Евк.»)<sup>1</sup>.

Как уже было отмечено выше, подробнее об этих вариантах, в частности – о способе их генерации, см. в [6], [7] и др.

Таблица 1. Результаты вычислительного эксперимента решения ЗКВ методом ветвей и границ

	случ.	псевдо, $\sigma = 1.4$	псевдо, $\sigma = 0.2$	Евк.
55	53	741	811	980
65	49	728	843	943
75	60	734	848	951
99	60	720	863	927

Для каждой клетки таблицы мы сгенерировали 1000 матриц для соответствующих ЗКВ (частных случаев проблемы). В таблице приведено число вариантов, при которых за первые 30 шагов фиксировался бы хоть 1 случай совпадения матриц – т.е. равенства 0 простейшей (дискретной) матричной метрики для некоторых двух из полученных за 30 шагов подзадач. Отметим, что получаемые результаты, по-видимому, не зависят (или почти не зависят) от размерности задач – но, как и ожидалось, имеют устойчивую тенденцию к увеличению при уменьшении значения  $\sigma$ . Однако, как мы уже отмечали, применять МВГ для геометрического (Евклидова) случая, т.е. при  $\sigma = 0$ , нецелесообразно.

### III. Краткое статистическое исследование задачи восстановления матриц ДНК малых размерностей

В настоящем разделе будет рассмотрена задача восстановления матрицы расстояний между последовательностями ДНК. Эта задача относится к области биокибернетики и ранее была описана авторами в [8], [9]. Задача заключается в восстановлении элементов матрицы расстояний между последовательностями ДНК. Как правило, при этом известно около 50% элементов матрицы. Для решения данной задачи целесообразно использовать т.н. *anupime*-алгоритм [7], который позволял бы отслеживать поэтапное восстановление элементов матрицы расстояний. В работе [10] исследуются алгоритмы решения задачи восстановления матрицы низкого ранга

<sup>1</sup>Очевидно, что Евклидов вариант может рассматриваться как псевдогеометрический с  $\sigma = 0$ . Кроме того заметим, что случайный вариант – при некоторых непринципиальных допущениях – также может быть рассмотрен как псевдо-геометрический, с «очень большим» значением  $\sigma$ .

с произвольно поврежденной долей ее элементов, эта проблема может рассматриваться как надежная версия классического PCA и возникает в ряде областей применения, включая обработку изображений, ранжирование Web-данных и биоинформационный анализ данных.

Алгоритм восстановления матрицы расстояний основан на анализе всех возможных треугольников, имеющих в матрице расстояний. Для каждого треугольника вычисляется значение *badness*, отслеживание которого является частью эвристического подхода (см. подробнее в [8]). Однако при проходе по элементам матрицы «слева направо» и «сверху вниз» можно получить нежелательные результаты с существенным увеличением *badness* для ранее исследованных треугольников. Одним из подходов для обхода данной проблемы является формирование последовательности подзадач с помощью модификации метода ветвей и границ [9]. В этом случае *badness* значение матрицы будет выполнять роль значения границы в классическом методе ветвей и границ [11], [12]. Нашей целью является рассмотрение различных статистических показателей, возникающих при восстановлении матриц описанным в [9] модифицированным методом ветвей и границ.

Ниже приведены результаты проведённого нами *статистического* исследования матрицы ДНК малых размерностей для задачи восстановления. Результаты проведённых нами вычислительных экспериментов приведены в следующей таблице (см. далее). Как и в случае предыдущей задачи (ЗКВ), мы отвечали только на один вопрос: *насколько часто встречается хотя бы 1 пара аналогичных матриц при применении нескольких первых шагов «классического» МВГ.* Однако, по сравнению с ЗКВ и предыдущим разделом, мы в этой формулировке заменили слово «одинаковых» на слово «аналогичных»: аналогичными мы в данном случае считаем такие матрицы, в которых совпадают множества незаполненных элементов. Несложно убедиться (также, например, путём вычислительных экспериментов), что дальнейшая работа МВГ с двумя аналогичными матрицами в подавляющем большинстве случаев происходят с одним и тем же порядком выбора ещё не заполненных элементов матрицы.

Генерация входных данных осуществлялась на основе матрицы, полученной алгоритмом Нидлмана – Вунша [13]. Мы применили этот алгоритм к цепочкам мДНК различных животных, взятых из банка данных NCBI [14]; при этом были взяты секвенированные цепочки мДНК для одного представителя каждого из 28 отрядов млекопитающих (классификацию млекопитающих выбираем согласно [15]). Для этой матрицы мы случайным образом, применяя равномерное распределение, выбирали нужное число её строк/столбцов (5, 10 или 15), получали матрицу меньшей размерности, в которой оставляли нужный процент элементов (от 30% до 70%). Для каждой пары, состоящей из размерности и процента удалённых элементов, мы делали 1000 подобных генераций. Далее мы запускали метод ветвей и границ – но, в отличие от рассмотрения ЗКВ в прошлом разделе, последним параметром было не число шагов МВГ, а число получающихся подзадач: при получении 10 (либо 30) подзадач мы вычисления прекращали. Также, конечно, мы прекращали вычисления и при получении двух аналогичных матриц – а именно количество таких вариантов (из 1000

возможных), что и отражено в таблице 2.

Таблица 2. Результаты вычислительного эксперимента восстановления матрицы методом ветвей и границ

	30	40	50	60	70
5	142/143	121/128	87/100		
10		196/401	173/363	166/334	
15			294/592	210/456	334/356

В ней:

- по строкам указана размерность;
- по столбцам указан процент удалённых элементов (так, при размерности 10 мы имеем всего 45 элементов, расположенных выше главной диагонали; удаление 40% означает удаление 18 элементов из них);
- в каждой из заполненных клеток – результаты вычислений для этого случая: число случаев (из 1000 рассмотренных), для которых получалась хотя бы одна пара;
- при этом первое значение в клетке приведено для вычислений, которые прекращались после получения 10 подзадач, а второе – после получения 30 подзадач.

#### IV. Применение метода ветвей и границ в задаче восстановления матрицы ДНК: результаты вычислительных экспериментов

У каждой матрицы есть ряд характеристик, влияющих на итог работы метода ветвей и границ. Одна из таких характеристик, входящая в этот список – это процент удалённых элементов. Ниже приводится (рис. 1) подборка результатов работы метода для задачи, рассмотренной в предыдущем разделе – для матриц размером  $5 \times 5$  при удалении 40%, 50% и 60% элементов исходной матрицы.

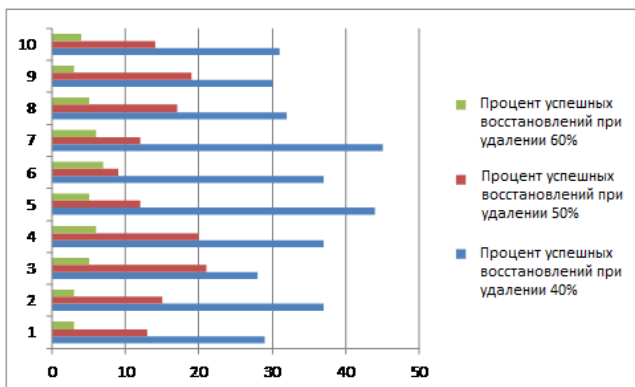


Рис. 1. Диаграмма распределения процента успешных восстановлений в зависимости от процента удалённых элементов для матриц размером  $5 \times 5$

На следующем графике (рис. 2) представлен средний процент успешных восстановлений в зависимости от процента удалённых элементов.

Далее для матриц размером  $10 \times 10$  проведены результаты аналогичных численных экспериментов (рис. 3).

Приведём также график средних показателей количества успешных восстановлений для матриц размером  $10 \times 10$  (рис. 4).

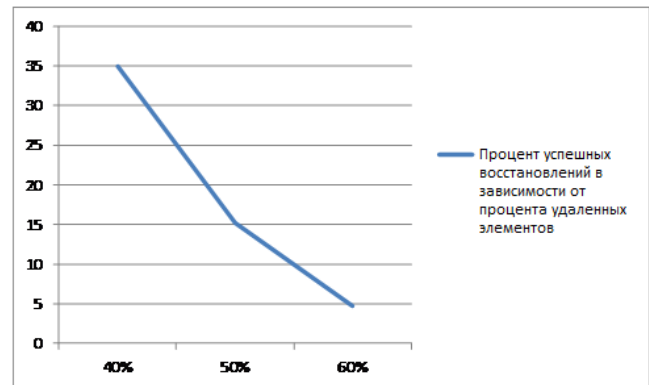


Рис. 2. График среднего процента успешных восстановлений в зависимости от процента удалённых элементов для матриц размером  $5 \times 5$ .

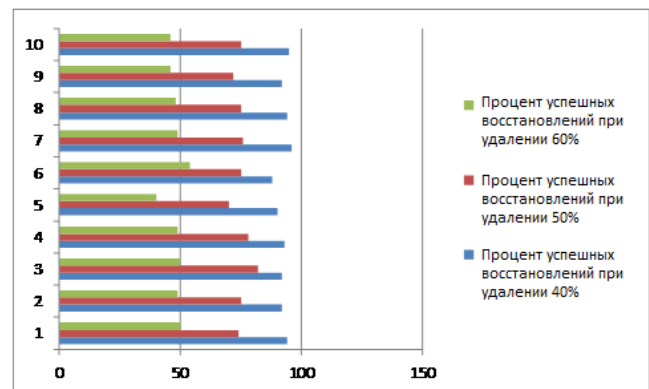


Рис. 3. Диаграмма распределения процента успешных восстановлений в зависимости от процента удалённых элементов для матриц размером  $10 \times 10$ .

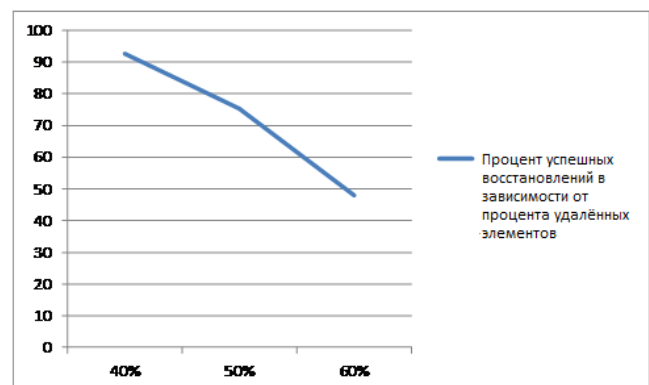


Рис. 4. График среднего процента успешных восстановлений в зависимости от процента удалённых элементов для матриц размером  $10 \times 10$ .

По сравнению с приведёнными ранее результатами для матриц  $5 \times 5$  заметно, что процент успешных восстановлений значительно увеличился. Это, по-видимому, связано с тем, что в ходе вычислительного процесса есть больше возможностей для выбора треугольников и восстановления элементов. При дальнейшем увеличении размерности матриц мы наблюдаем стремление процента успешных восстановлений к 100 при указанных процентах удалённых элементов (40%, 50%, 60%) (рис. 5).

Следующей интересной характеристикой является высота дерева решений (т.е. максимальная длина пути от корня дерева к листу). На представленных диаграммах (рис. 6 и 7) хорошо видно, что среднее значение высоты

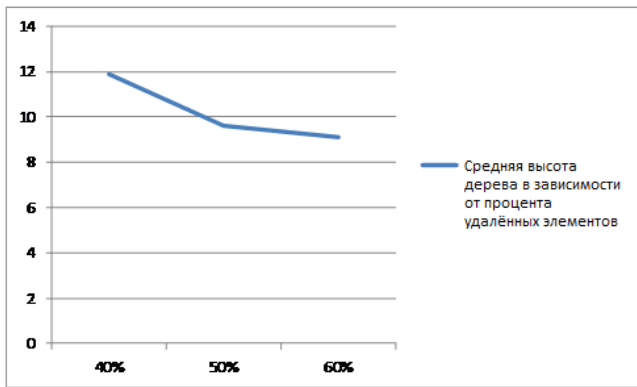


Рис. 5. Высота дерева решений в зависимости от процента удалённых элементов.

дерева решений для матрицы  $5 \times 5$  не слишком изменяется в зависимости от процента удалённых элементов.

Имеет также смысл посмотреть изменение высоты дерева решений в зависимости от размерности матрицы при фиксированном количестве удалённых элементов. Две представленные далее диаграммы иллюстрируют это сравнение.

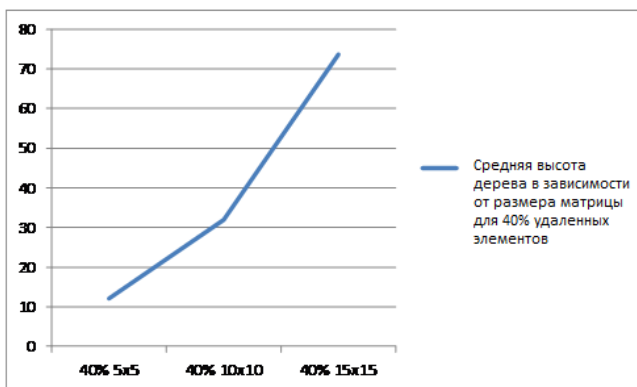


Рис. 6. Высота дерева решений в зависимости от размера матрицы для 40% удалённых элементов.

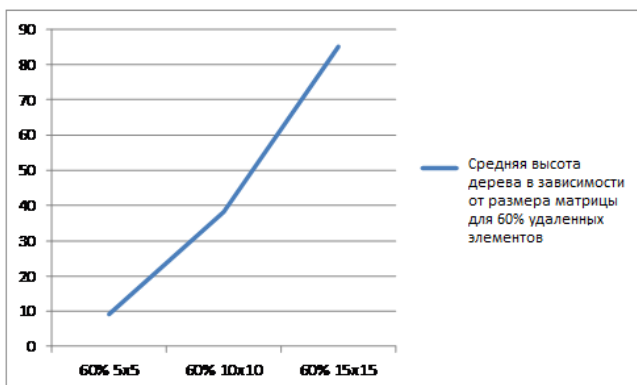


Рис. 7. Высота дерева решений в зависимости от размера матрицы для 60% удалённых элементов.

Можно видеть, что увеличение высоты дерева решений при росте размерности матрицы происходит с высокой скоростью. Также при увеличении размерности матрицы существенно возрастает количество подзадач одного уровня, и ввиду этого перебор задач занимает значительно большее время.

## V. Краткое рассмотрение примера малой размерности псевдогеометрической версии задачи коммивояжера

В этом разделе мы рассмотрим пример малой размерности. Причём сначала мы покажем, как именно, используя датчик случайных чисел, получается геометрическая версия ЗКВ (именно такие частные случаи ЗКВ рассматриваются в качестве примеров геометрических версий в [5], [16] и статьях, процитированных в этих монографиях). Далее приведём пример восстановления точек по матрице, полученной по этой версии – как и следовало ожидать, восстановленное множество точек совпадает с исходным с точностью до поворота и параллельного переноса. Далее мы с помощью того же самого датчика случайных чисел получим вариант псевдогеометрической ЗКВ – и после т.н. «псевдовосстановления» (т.е. применения к полученным данным псевдогеометрической ЗКВ алгоритма, совпадающего с алгоритмом восстановления для геометрической версии, см. подробнее в [6], [17], [18]) получаем приемлемое значение невязки – т.е. вычисленного по некоторой естественной матричной метрике расстояния между исходной матрицей геометрической ЗКВ и полученной матрицей.

Таблица 3. Координаты 8 случайных выбранных с помощью равномерного распределения на отрезке  $[0, 1]$  точек

	X	Y
1	0.779	0.76
2	0.274	0.157
3	0.601	0.952
4	0.919	0.45
5	0.533	0.21
6	0.49	0.758
7	0.024	0.457
8	0.149	0.92

В таблице 3 мы привели координаты случайно заданных (выбранных с помощью равномерного распределения на отрезке  $[0, 1]$ ) 8 точек<sup>2</sup> – по которым сначала построим матрицу ЗКВ размерности  $8 \times 8$ . При этом важно отметить, что в алгоритмах, связанных с решением частных случаев ЗКВ, мы не знаем координаты исходных точек. Построенная по этим точкам матрица такова (таблица 4):

Таблица 4. Матрица ЗКВ размерности 8, построенная по выбранным точкам

	1	2	3	4	5	6	7	8
1	99.999	0.786	0.262	0.34	0.602	0.29	0.814	0.651
2	0.786	99.999	0.86	0.708	0.264	0.638	0.391	0.773
3	0.262	0.86	99.999	0.594	0.745	0.224	0.761	0.454
4	0.34	0.708	0.594	99.999	0.454	0.528	0.895	0.902
5	0.602	0.264	0.745	0.454	99.999	0.549	0.566	0.807
6	0.29	0.638	0.224	0.528	0.549	99.999	0.555	0.378
7	0.814	0.391	0.761	0.895	0.566	0.555	99.999	0.48
8	0.651	0.773	0.454	0.902	0.807	0.378	0.48	99.999

<sup>2</sup>В отличие от остальных разделов (где использовались сложные вычислительные эксперименты, реализованные нами варианты метода ветвей и границ и т.п.), все результаты настоящего раздела можно легко проверить с помощью простых вспомогательных программ (написанных на Excel или на каком-либо языке программирования) – или даже «вручную».

Применив на её основе алгоритм восстановления, мы получаем следующие 8 точек (таблица 5):

Таблица 5. Координаты 8 точек, полученных в результате применения алгоритма восстановления

	X	Y
1	0.000	0.000
2	0.000	0.786
3	-0.260	-0.034
4	0.306	0.148
5	0.164	0.579
6	-0.221	0.187
7	-0.384	0.717
8	-0.586	0.282

а на их основе можно получить матрицу  $8 \times 8$ , которая окажется (как и следовало ожидать) в точности совпадающей с приведённой выше.

По-видимому, представляет интерес рисунок с расположением этих точек; на приведённом рисунке исходные точки показаны синим цветом, а полученные – зелёным (рис. 8):

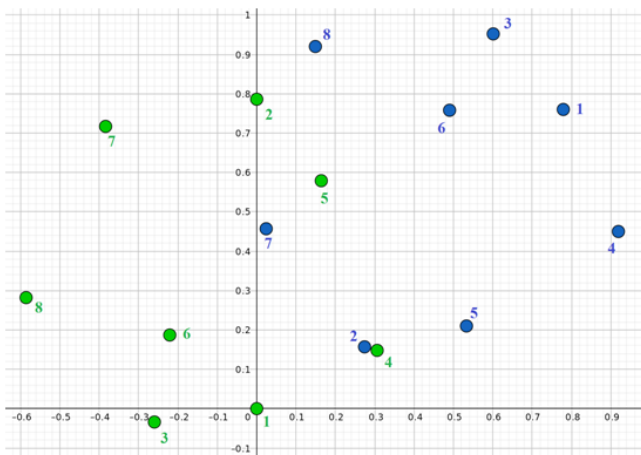


Рис. 8. Расположение на координатной плоскости 8 точек, выбранных случайным образом и полученных с помощью алгоритма восстановления.

Однако, конечно, существенно более важными являются дальнейшие построения. Применяя тот же самый датчик случайных чисел, мы на основе предыдущей матрицы  $8 \times 8$  получаем следующую исходную матрицу псевдо-геометрической (таблица 6) ЗКВ<sup>3</sup>:

Таблица 6. Матрица псевдо-геометрической ЗКВ размерности  $8 \times 8$

	1	2	3	4	5	6	7	8
1	99.999	0.994	0.285	0.294	0.662	0.313	0.786	0.647
2	0.994	99.999	0.758	0.735	0.219	0.716	0.368	0.51
3	0.285	0.758	99.999	0.598	0.795	0.216	0.907	0.379
4	0.294	0.735	0.598	99.999	0.513	0.536	0.871	0.979
5	0.662	0.219	0.795	0.513	99.999	0.528	0.441	0.836
6	0.313	0.716	0.216	0.536	0.528	99.999	0.648	0.446
7	0.786	0.368	0.907	0.871	0.441	0.648	99.999	0.552
8	0.647	0.51	0.379	0.979	0.836	0.446	0.552	99.999

<sup>3</sup>По-видимому, конкретные параметры датчика и конкретные версии применяемых нами алгоритмов не представляют особого интереса. Отметим только, что нами применялось нормальное распределение с  $\mu = 1$  (что является единственно возможным вариантом) и дисперсией  $5/3$ .

К полученной матрице мы применяем алгоритм, аналогичный алгоритму восстановления (как мы уже отметили, некоторые подробности см. в [5], [16], [19]) – после чего получаем следующее множество точек (таблица 7):

Таблица 7. Координаты 8 точек, полученных в результате применения алгоритма восстановления псевдо-геометрической ЗКВ

	X	Y
1	0.000	0.000
2	0.000	0.994
3	0.138	0.249
4	-0.161	0.337
5	-0.245	0.692
6	-0.004	0.274
7	-0.144	0.651
8	0.383	0.694

По этому множеству мы получаем такую матрицу  $8 \times 8$  (таблица 8):

Таблица 8. Матрица ЗКВ размерности 8, полученная по восстановленным точкам псевдо-геометрической ЗКВ

	1	2	3	4	5	6	7	8
1	99.999	0.994	0.285	0.374	0.734	0.274	0.666	0.793
2	0.994	99.999	0.758	0.676	0.39	0.72	0.373	0.487
3	0.285	0.758	99.999	0.311	0.585	0.144	0.49	0.508
4	0.374	0.676	0.311	99.999	0.364	0.169	0.314	0.65
5	0.734	0.39	0.585	0.364	99.999	0.482	0.11	0.628
6	0.274	0.72	0.144	0.169	0.482	99.999	0.401	0.571
7	0.666	0.373	0.49	0.314	0.11	0.401	99.999	0.528
8	0.793	0.487	0.508	0.65	0.628	0.571	0.528	99.999

– и стоит отметить, что некоторая «естественная» численная характеристика (метрика), показывающая отличие этой матрицы от исходной матрицы псевдо-геометрической ЗКВ (т.е. от предыдущей матрицы), даёт вполне приемлемые результаты<sup>4</sup>.

## VI. Заключение

Полученные нами в настоящей статье (прежде всего – в разделах II и III) статистические закономерности фактически являются вероятностью возникновения удачной ситуации – которая даёт возможность не вычислять разделяющий элемент лишней раз. (По нашим подсчётам, полученным также в ходе вычислительных экспериментов, на выбор разделяющего элемента в некоторых вариантах МВГ тратится более 99% времени работы программы.) Поэтому полученные результаты дают обоснование применения кластеризации ситуаций при разработке алгоритмов решения задач дискретной оптимизации с помощью метода ветвей и границ. Например, для ЗКВ такое применение даёт легко наблюдаемые улучшения работы алгоритма прежде всего для псевдо-геометрического варианта; именно этот вариант, с нашей точки зрения, отражает репрезентативность данных во многих реальных задачах.

<sup>4</sup>Подробности опять же приводить не будем. Если совсем кратко – мы использовали «нормированный квадратный корень» из суммы квадратов разностей соответствующих элементов; эта метрика в данном случае даёт приемлемое значение 0.208.

## Список литературы

- [1] Макаркин С. Б., Мельников Б. Ф., Тренина М. А. Подход к решению псевдогеометрической версии задачи коммивояжера // Известия высших учебных заведений. Поволжский регион. Физико-математические науки. – 2015.– № 2 (34). – С. 135–147.
- [2] Мельников Б.Ф. Мельникова Е. А. Некоторые эвристические алгоритмы в задаче вершинной минимизации недетерминированных конечных автоматов // Стохастическая оптимизация в информатике.– 2013. –Т. 9, № 2. – С. 73–87.
- [3] Мельников Б. Ф., Мельникова Е. А. Подход к программированию недетерминированных игр (Часть I: Описание общих эвристик) // Известия высших учебных заведений. Поволжский регион. Физико-математические науки. – 2013/ – № 4 (28). – С. 29–38.
- [4] Мельников Б. Ф., Давыдова Е. В., Ничирпочук А. В., Тренина М. А. Кластеризация ситуаций в алгоритмах решения задачи коммивояжера и её применение в некоторых прикладных задачах. Часть II. Списочная метрика и некоторые связанные оптимизационные проблемы // Известия высших учебных заведений. Поволжский регион. Физико-математические науки. – 2018.– № 4. [https://izvuz\\_fm.npznzgu.ru/fmn418](https://izvuz_fm.npznzgu.ru/fmn418).
- [5] Gutin G., Punnen A. P. Combinatorial Optimization. The Traveling Salesman Problem and Its Variations. – Berlin: Springer, 2007. – 830 p.
- [6] Макаркин С. Б., Мельников Б. Ф. Геометрические методы решения псевдогеометрической версии задачи коммивояжера // Стохастическая оптимизация в информатике. – 2013. – Т. 9, № 2.– С. 54–72.
- [7] Melnikov V. Multiheuristic approach to discrete optimization problems // Cybernetics and Systems Analysis. – 2006. – Vol. 42, No. 3. – P. 335–341.
- [8] Мельников Б. Ф., Тренина М. А. Об одной задаче восстановления матриц расстояний между цепочками ДНК // International Journal of Open Information Technologies. – 2018. – Vol. 6, No. 6.– С. 1–13.
- [9] Мельников Б. Ф., Тренина М. А. Применение метода ветвей и границ в задаче восстановления матрицы расстояний между цепочками ДНК// International Journal of Open Information Technologies. – 2018. – Vol. 6, No. 8. – С. 1–13.
- [10] Ganesh A., Lin Z., Wright J., Wu L., Chen M. Fast algorithms for recovering a corrupted low-rank matrix // Computational Advances in Multi-Sensor Adaptive Processing (CAMSAP). 2009. – 3rd IEEE International Workshop on 2009. – pp.213–216.
- [11] Гудман С, Хидетгними С. Введение в разработку и анализ алгоритмов. – М.: Мир, 1981. – 364 с.
- [12] Hromkovic J. Algorithmics for Hard Problems. Introduction to Combinatorial Optimization, Randomization, Approximation, and Heuristics. – Basel: Springer, 2004. – 536 p.
- [13] Мельников Б. Ф., Тренина М. А., Кочергин А. С. Подход к улучшению алгоритмов расчёта расстояний между цепочками ДНК (на примере алгоритма Нидлмана – Вунша) // Известия высших учебных заведений. Поволжский регион. Физико-математические науки. – 2018. – № 1 (45). – С. 46–59.
- [14] Home – Nucleotide – NCBI [Электрон. ресурс]. – Режим доступа: <https://www.ncbi.nlm.nih.gov/nucleotide>, свободный
- [15] Айала Ф, Кайгер Дж. Современная генетика. Пер. с англ. Т. 1. – М.: Мир, 1987. – 295 с.
- [16] Applegate D. L., Vixby R. E., Chvatal V., Cook W. J. The Traveling Salesman Problem: A Computational Study. – NJ: Princeton University Press, 2007.– 593 p.
- [17] Мельников Б. Ф., Тренина М. А., Макаркин С. Б. Применение проблемно-ориентированных метрик в геометрических алгоритмах решения псевдогеометрической версии задачи коммивояжера // Стохастическая оптимизация в информатике. – 2014. – Т.10, № 1. – С. 63–71.
- [18] Мельников Б. Ф., Давыдова Е. В. Математическое моделирование повышения уровня безопасности в случае отказов авиационной и космической техники // International Journal of Open Information Technologies. – 2018. – Vol. 6, No. 5. – P. 1–6.
- [19] Гимади Э. Х., Хачай М. Ю. Экстремальные задачи на множествах перестановок. – Екатеринбург: Издательство УМЦ УПИ, 2016. – 220 с.

Борис Феликсович Мельников,  
 профессор Российского государственного социального университета (<http://www.rgsu.net/>),  
 профессор Московского государственного университета им. М. В. Ломоносова (<https://www.msu.ru/>),  
 email: [bf-melnikov@yandex.ru](mailto:bf-melnikov@yandex.ru),  
[mathnet.ru](mailto:mathnet.ru): personid=27967,  
[elibrary.ru](mailto:elibrary.ru): authorid=15715,  
[scopus.com](https://scopus.com): authorId=55954040300.

Анастасия Владимировна Ничипорчук,  
 аспирант Российского государственного социального университета (<http://www.rgsu.net/>),  
 email: [anichiporчук@gmail.com](mailto:anichiporчук@gmail.com),  
[elibrary.ru](mailto:elibrary.ru): authorid=1001250.

Марина Анатольевна Тренина,  
 старший преподаватель Тольяттинского государственного университета (<https://www.tltsu.ru/>),  
 email: [trenina.m.a@yandex.ru](mailto:trenina.m.a@yandex.ru),  
[elibrary.ru](mailto:elibrary.ru): authorid=235496,  
[scopus.com](https://scopus.com): authorId=57204505140.

Михаил Эдуардович Абрамян,  
 доцент Южного федерального университета (<http://http://sfedu.ru/>)  
 email: [m-abramyan@yandex.ru](mailto:m-abramyan@yandex.ru),  
 (<http://sfedu.ru>),  
 email: [m-abramyan@yandex.ru](mailto:m-abramyan@yandex.ru),  
[mathnet.ru](mailto:mathnet.ru): personid=20226,  
[elibrary.ru](mailto:elibrary.ru): authorid=17911,  
[scopus.com](https://scopus.com): authorId=8291167000.



# Clustering of situations in solving applied optimization problems (on the examples of traveling salesman problem and distance matrix recovery)

Boris Melnikov, Anastasia Nichiporchuk, Marina Trenina, Mikhail Abramyan

**Abstract**—In discrete optimization problems, we apply algorithms based on extensions of the branch and bound method. These extensions consist in the joint work of several auxiliary heuristic algorithms, they can be referred to different, independent from each other, areas of artificial intelligence. Therefore, the relevance of the problems under consideration is provided by both subject areas and algorithms.

In this paper, we investigate the possibility of using one of these auxiliary algorithms, so-called clustering of situations. As the subject areas, we consider two different discrete optimization problems: the traveling salesman problem in its classical formulation (we prefer to study its special cases obtained for the pseudogeometric version) and the problem of DNA distance matrix reconstruction. As a result of computational experiments, we obtained some regularities that allow us to create improved versions of the branch and bound algorithm – by connecting heuristics to it for clustering situations.

The results obtained in computational experiments provide a rationale for the application of clustering situations in the development of algorithms using the branch and bound method. For example, for the traveling salesman problem, this application gives easily observable improvements in the algorithm, primarily for the pseudogeometric version.

**Keywords**—heuristic algorithms, discrete optimization problems, branch and bound method, DNA distance matrix, situation clustering.

## REFERENCES

- [1] Makarkin S. B., Melnikov B.F., Trenina M.A. The approach to solving the pseudo-geometric version of the traveling salesman problem // Proceedings of Higher Educational Institutions. Volga Region. Physical and Mathematical Science. – 2015.– № 2 (34). – C. 135–147. (in Russian)
- [2] Melnikov B.F., Melnikov E.A. Some heuristic algorithms in the vertex minimization problem of nondeterministic finite automata // Stochastic optimization in computer science.– 2013. –T. 9, No 2. – C. 73–87. (in Russian)
- [3] Melnikov B.F., Melnikov E.A. Approach to programming nondeterministic GAMES (Part I: Description of general heuristics) // Proceedings of higher educational institutions. Volga Region. Physics and mathematics. – 2013/ – № 4 (28). – C. 29–38. (in Russian)
- [4] Melnikov B.F., Davydova E.V., Nichiporchuk A.V., Trenina M.A. Clustering of situations in algorithms for solving the traveling salesman problem and its application in some applications. Part II. List metric and some related optimization problems // Proceedings of Higher Educational Institutions. Volga Region. Physical and Mathematical Science. – 2018.– № 4. [https://izvuz\\_fm.n.pnzgu.ru/fmn418](https://izvuz_fm.n.pnzgu.ru/fmn418). (in Russian)
- [5] Gutin G., Punnen A.P. Combinatorial Optimization. The Traveling Salesman Problem and Its Variations. – Berlin: Springer, 2007. – 830 p.
- [6] Makarkin S. B., Melnikov B.F. Geometric methods for solving the pseudo-geometric version of the traveling salesman problem // Stochastic optimization in computer science. – 2013. – T. 9, № 2.– C. 54–72. (in Russian)
- [7] Melnikov B. Multiheuristic approach to discrete optimization problems // Cybernetics and Systems Analysis. – 2006. – Vol. 42, No. 3. – P. 335–341.
- [8] Melnikov B.F., Trenina M.A. On a problem of the reconstruction of distance matrices between DNA sequences // International Journal of Open Information Technologies. – 2018. – Vol. 6, No. 6.– C. 1–13. (in Russian)
- [9] Melnikov B.F., Trenina M.A. The application of the branch and bound method in the problem of reconstructing the matrix of distances between DNA strings // International Journal of Open Information Technologies. – 2018. – Vol. 6, No. 8. – C. 1–13. (in Russian)
- [10] Ganesh A., Lin Z., Wright J., Wu L., Chen M. Fast algorithms for recovering a corrupted low-rank matrix // Computational Advances in Multi-Sensor Adaptive Processing (CAMSAP). 2009. – 3rd IEEE International Workshop on 2009. – pp. 213–216.
- [11] Goodman, S. Introduction to the Design and Analysis of Algorithms / S. Goodman, S. Hedetniemi – NY: McGraw-Hill; 1977. – 344 p.
- [12] Hromkovic J. Algorithmics for Hard Problems. Introduction to Combinatorial Optimization, Randomization, Approximation, and Heuristics. – Basel: Springer, 2004. – 536 p.
- [13] Melnikov B.F., Trenina M.A., Kochergin A.S. An approach to improving algorithms for calculating distances between DNA strings (using the example of the Needleman-Wunsch algorithm)// Proceedings of higher educational institutions. Volga Region. Physics and mathematics. – 2018. – № 1 (45). – C. 46–59. (in Russian)
- [14] Home – Nucleotide – NCBI [Электрон. ресурс]. – Режим доступа: <https://www.ncbi.nlm.nih.gov/nucleotide>, свободный
- [15] Ayala, F. J., Kiger, J.F. (1980), Modern Genetics, 2nd ed., Menlo Park: University of California at Danis. – 724 p.

- [16] Applegate D.L., Bixby R.E., Chvatal V., Cook W.J. The Traveling Salesman Problem: A Computational Study. – NJ: Princeton University Press, 2007.– 593 p.
- [17] Melnikov B.F., Trenina M.A., Makarkin S. B. Using Problem Oriented Metrics for Solving Pseudo-Euclidian Travelling Salesman Problem with Euclidian Algorithms // Stochastic optimization in computer science. – 2014. – T.10, № 1. – С. 63–71. (in Russian)
- [18] Melnikov B. F., Davydova E. V. Mathematical modeling of increasing the level of safety in case of failures of space technology // International Journal of Open Information Technologies. – 2018. – Vol. 6, No. 5. – P. 1–6. (in Russian)
- [19] Gimadi E.H., Khachay M. Yu. Extremal problems on sets of permutations. – Yekaterinburg: publishing house of the UMTS UPI, 2016. — 220 p. (in Russian)