

**УТОЧНЕНИЕ РЕШЕНИЙ АЛГОРИТМОВ ЖАДНОГО  
ПОИСКА КРАТЧАЙШЕГО НЕЗАМКНУТОГО ПУТИ  
МЕЖДУ ОБЪЕКТАМИ ДЛЯ ВИЗУАЛИЗАЦИИ  
МНОГОМЕРНЫХ ДАННЫХ**

***Сурков Егор Эдуардович***

*магистрант, каф. ИБ*

*Федеральное государственное бюджетное образовательное учреждение  
высшего образования «Тульский государственный университет», кафедра  
информационной безопасности, Тула, Россия*

*E-mail: eg-su@mail.ru*

***Научный руководитель — Середин Олег Сергеевич***

Одной из целей визуализации данных является упрощение их восприятия для последующего анализа. Визуализировать данные, признаковое пространство которых больше трех, проблематично, что вызвало появление методов решения этой задачи [1]. Основной идеей визуализации многомерных данных в нашей предыдущей работе [1] является визуализация на основе поиска кратчайшего незамкнутого пути между объектами (точками) исследуемой выборки и его отображением на двумерную плоскость в виде незамкнутого графа (цепочки), столбчатой диаграммы распределения объектов вдоль найденного пути или проекции на путь [1]. Кратчайший незамкнутый путь (КНП) – это граф из  $N - 1$  ребер, который соединяет  $N$  объектов пространства таким образом, чтобы суммарная длина ребра была минимальной [1]. В работе предлагается к рассмотрению несколько критериев поиска КНП. 1. Первый критерий заключается в поиске кратчайшего незамкнутого пути и математически выражается как минимизация следующего функционала:

$$J_1 = \min \sum_{i=2}^N d_{i,i-1}, \quad (1)$$

где  $d_{i,j}$  - расстояние между  $i$ -й и  $j$ -й точкой,  $i = 1, \dots, N$  - порядок обхода элементов в пути,  $N$  - количество элементов. 2. Второй критерий обеспечивает минимум разницы между длиной КНП и расстоянием между терминальными точками пути:

$$J_2 = \min \left( \sum_{i=2}^N d_{i,i-1} - \alpha \cdot d_{1,N} \right), \quad (2)$$

где  $\alpha$  - весовой коэффициент (в работе используется  $\alpha = 1$ ). Для

поиска кратчайшего незамкнутого пути в работе [1] предложены 9 алгоритмов. Алгоритм A0, который заключается в полном переборе всех возможных комбинаций точек и выборе пути, сумма расстояний между точками которого является минимальной. А также предложены различные модификации жадных алгоритмов A1-A4 и их рекурсивные версии A1R-A4R. Идея алгоритмов заключается в следующем: определяется некоторое количество терминальных точек, из которых строится КНП таким образом, что на каждом шаге для каждой точки находится ближайшая, и так до тех пор, пока не будут соединены все точки. В зависимости от модификации изменяется правило выбора затравочных точек, их количество, временные затраты на вычисление и пространство решений алгоритмов [1]. Экспериментальные исследования продемонстрировали, что алгоритмы A1-A4R успешно справляются с решением задачи КНП на небольших данных [1], а также с эффективностью модификаций жадных алгоритмов растет и их вычислительная сложность. Строго говоря, кратчайший незамкнутый путь может быть найден только алгоритмом A0, а остальные алгоритмы выполняют поиск квази-кратчайшего незамкнутого пути. В этой работе рассматривается идея улучшения решений, полученных алгоритмами A1-A4R. Для улучшения решений, полученных при помощи модификаций жадного алгоритма, в работе предлагается использовать муравьиный алгоритм [5], который позволит осуществить поиск кратчайшего незамкнутого пути между двумя точками. Этими точками будут являться терминальные точки из КНП, найденных алгоритмом A4R [1]. «Муравьиный» алгоритм моделирует процесс естественного поведения муравьев в природе и основан на исследовании прилегающей к муравейнику территории (начальная точка) на наличие источников питания (конечная точка), помечая более удачные пути от колонии до источника большим количеством феромона [5]. Принцип работы алгоритма заключается в следующем: виртуальный «муравей» помещается в начальную вершину. Исходя из данных о длине ребра и количестве феромона, отложенного на нем предыдущими поколениями «муравьев», определяется вероятность перехода в следующие доступные точки. Начальное количество феромона на всех ребрах инициализируется стартовым значением, отличным от нуля. Стоит отметить, что оригинальная идея алгоритма состоит в поиске кратчайшего пути от начальной точки до конечной без учета обязательного прохождения через все точки. Для решения задачи поиска кратчайшего пути через все точки необходимо ввести очередь,

которая содержит все еще непосещённые вершины за исключением конечной. К моменту, когда все вершины из очереди посещены, проверяется доступность перехода из последней в пути точки в терминальную. Если переход доступен, то путь считается успешным.

В заключительной части работы приведены результаты экспериментов на модельных и реальных данных, таких как Iris Data Set [2], Abalone Data Set [3], а также на базисной совокупности скелетных моделей [4]. Выполнено экспериментальное сравнение алгоритмов поиска кратчайшего незамкнутого пути, а также приведены сравнительные таблицы временных затрат на их вычисление. Результат, полученный базовыми алгоритмами, удалось улучшить, применив муравьиный алгоритм [5]. Например, на Ирисах Фишера [2] получилось построить КНП, длина которого сократилась с 51.627 до 50.898, а на данных из задачи детектирования падений [4] длина пути уменьшилась с 5.95 до 5.897.

Работа выполнена при финансовой поддержке Министерства науки и высшего образования РФ в рамках государственного задания FEWG-2021-0012, гранта ректора ТулГУ 8925ГРР\_М.

### Литература

1. Середин О. С., Сурков Е. Э., Копылов А. В., Двоенко С. Д. Визуализация многомерных данных на основе построения кратчайшего незамкнутого пути // Известия ТулГУ. Технические науки. 2021. №12, с. 299-312.
2. Ирисы Фишера. [Электронный ресурс] / – Режим доступа: [https://ru.wikipedia.org/wiki/Ирисы\\_Фишера](https://ru.wikipedia.org/wiki/Ирисы_Фишера). – Дата доступа: 15.05.2021.
3. Dua D. and Graff C. (2019). UCI Machine Learning Repository [<http://archive.ics.uci.edu/ml>]. Irvine, CA: University of California, School of Information and Computer Science.
4. Сурков Е. Э. Исследование базисной совокупности скелетных представлений в задаче детектирования падений // Ломоносов-2021 : Сборник тезисов XXVIII Международной научной конференции студентов, аспирантов и молодых ученых, Москва, 12–23 апреля 2021 года / Сост. Е.И. Атамась, А.В. Мальцева. – Москва: Издательский отдел факультета ВМК МГУ, ООО "МАКС Пресс", 2021. – С. 82-83.
5. Кажаров А. А., Курейчик, В. М. Об одном «муравьином» алгоритме. [Электронный ресурс] – Режим доступа: <https://disk.yandex.ru/i/C8X1BAm2dglMLQ>. – Дата доступа: 05.12.2021.